Electroweak Gauge Bosons at Future Electron-Positron Colliders

Thorsten Ohl

Darmstadt University of Technology
Schloßgartenstr. 9
D-64289 Darmstadt, Germany

<ohl@hep.tu-darmstadt.de>

IKDA 99/11
June 1999

Abstract

The interactions of electroweak gauge bosons are severely constrained by the symmetries inferred from low energy observables. The exploration of the dynamics of electroweak symmetry breaking requires therefore both an experimental sensitivity and a theoretical precision that is better than the natural scale of deviations from minimal extrapolations of the low energy effective theory. Some techniques for obtaining improved theoretical predictions for precision tests at a future e^+e^- linear collider are discussed.

Preface

Several excellent, up-to-date reviews of the wide range of physics that can be probed at a future e^+e^- Linear Collider (LC) are available [1, 2, 3]. The present review discusses some physical and technical aspects of electroweak gauge bosons, that will be important at a LC. The importance of gauge invariance even in a conservative Effective Field Theory (EFT) approach, which does not pretend to be valid to arbitrarily short distances, is emphasized. The implementation of gauge invariance in perturbation theory is discussed. Techniques for improving the convergence and speed of high dimensional phase space integrations are described.

Acknowledgments

I thank Panagiotis Manakos, Hans Dahmen, and Thomas Mannel for their support over the years. Howard Georgi deserves the credit for the most inspirational year of physics so far (Howard provided an open door, physics insight, and inspiration, while DFG provided the financial support).

I thank Harald Anlauf, Edward Boos, Wolfgang Kilian, and Alex Nippe for enjoyable and productive collaborations that have contributed to some of the results presented in the following pages.

I thank the organizers of the Maria Laach school for the invitation to climb on the Effective Field Theory soapbox, as reproduced in part in appendix A. The students make this a fun place to lecture and the hospitality of the congregation of Maria Laach makes this school unique.

I thank Deutsche Forschungsgemeinschaft (DFG) and Bundesministerium for Bildung, Wissenschaft, Forschung und Technologie (BMBF) for financial support.

Contents

—1—
Introduction

1.1 The Art and Craft of Elementary Particle Physics

To teach how to live without certainty, and yet without being paralyzed by hesitation, is perhaps the chief thing that philosophy, in our age, can still do for those who study it.
[Bertrand Russell: *A History of Western Philosophy*]

In a highly idealized scenario, theoretical elementary particle physics proceeds by condensing the results of experimental observations up to some energy scale into a model, which is designed to remain consistent at higher energies. In a second step, predictions for experiments at these higher energy scales are derived. Finally, comparisons with subsequent experimental results are used to refine the model, if necessary, and the process starts anew.

Therefore, we can identify two complementary tasks: *model building* on one hand and *calculation of observables* on the other. Below, we will focus our attention in the area of model building to the systematic procedure of EFT and its rôle in the modern interpretation of the Standard Model (SM) with simple extensions. The methods for the calculation of observables will be discussed in greater detail. Such calculations are naturally divided into two separate parts: *calculation of amplitudes and differential cross sections* and *phase space integration*. Some methods for both parts are described with a noticeable bias towards the author's research interests.

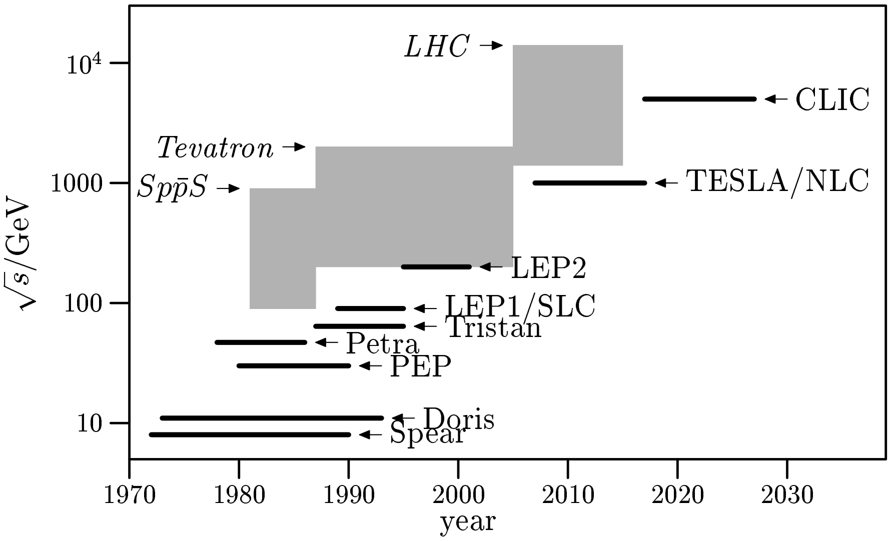


Figure 1.1: *Physics reaches of past, present and future colliders. Hadron colliders can not utilize the full beam energy in partonic collisions since the momentum is shared by quarks and gluons. As a rule of thumb, the lower border of the grey rectangles ($\sqrt{s'} = \sqrt{s}/10$) corresponds to the physics reach of hadron colliders. (Data taken from [8, 9].)*

1.2 Success and Failure of the Standard Model

Tout comprendre,
c'est tout mepriser.
[FRIEDRICH NIETZSCHE]

In the decade after the commissioning of LEP1, many aspects of the Standard Model [4] of elementary particle physics have been tested successfully to a precision previously unimaginable [5, 6]. Indeed it is not without irony that one of the main sources of theoretical uncertainty derives from the insufficient precision of the old data on hadron production in e^+e^--collisions at a few GeV, despite recent theoretical progress in the interpretation of this data [7].

The important qualitative features [10] of the SM were established from low energy experiments at a time before any of the colliders shown in figure 1.1 were commissioned. The symmetry structure of the SM as a $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge theory that is spontaneously broken to the U(1) of Quantum Electrodynamics (QED) has been firmly established by the stunning success

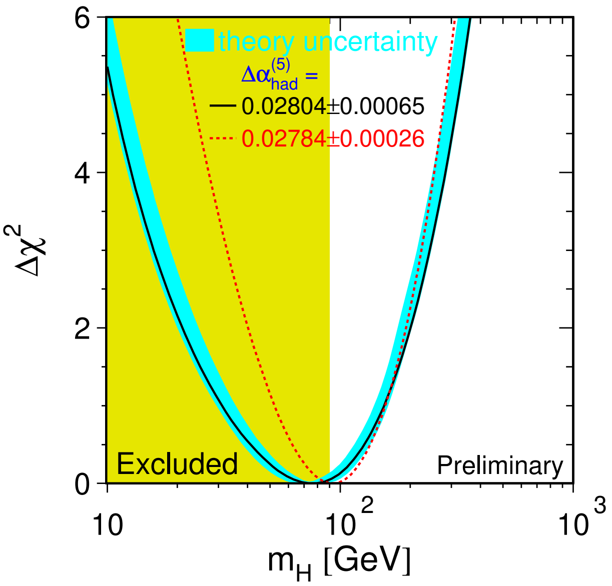
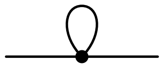


Figure 1.2: *The famous "blue band" derived from fits of electroweak precision data to SM predictions [6], that suggests a very light SM Higgs particle, with a mass close to the present exclusion limit. The area excluded by direct searches is shaded in light grey on the left. The parabolas show the distance in likelihood from the best fit as a function of the Higgs mass for two different determinations of the running electromagnetic coupling* $\alpha_{QED}(M_Z)$.

of the SM at LEP1. Nevertheless, direct information regarding the underlying dynamics of the Electro Weak Symmetry Breaking (EWSB) has yet to be found.

The existing data is compatible with the minimal SM with a light elementary Higgs particle [6]. Global fits of the SM predictions to the LEP1 precision observables (including $m_{\rm top}$ from Tevatron and M_W from LEP2) suggest a Higgs mass close to the present exclusion limits (see figure 1.2). However, the Higgs mass enters the precision observables only logarithmically and the limits are therefore weak.

Despite the phenomenological hints in its favor, it must be mentioned that the minimal SM with an elementary Higgs particle is unattractive from a theoretical point of view. There is a quadratically divergent contribution to the Higgs mass renormalization

$$\delta m^2_{\text{Higgs}} =$$  $$\propto \frac{g}{16\pi^2}\Lambda^2 \qquad (1.1)$$

that requires an unrealistic fine-tuning of the mass to one part in 10^{38} in order to keep the Higgs light and the EWSB scale close to 1 TeV (see section A.2.5 for more details).

There are two ways around this problem and both feature fermions prominently. A supersymmetry can relate the scalar mass to a fermion mass. Since a fermion mass is protected from renormalization by chiral symmetry, it can be naturally light and the scalar mass can naturally remain on the order of the supersymmetry breaking scale. On the other hand, Goldstone bosons from the spontaneous breaking of a continuous symmetry are protected by Goldstone's theorem and can be naturally light on the order of the symmetry breaking scale as well. Here the chiral symmetry of fermions is required to protect the particles that take part in the interaction, which is responsible for the symmetry breaking, from a large mass renormalization. The latter models are disfavored by current precision data.

A necessary, though not sufficient, condition for the unification of gauge interactions in a single simple group at a Grand Unified Theory (GUT) scale is that the running couplings (see sections A.2.8 and A.3.2 for details) have a common value at one scale. According to the data from the pre-LEP1 era, the SM has this property. However, LEP1 precision data revealed that the coupling do not meet, unless supersymmetric partners are added to the particle spectrum [11].

Even if the naturalness problem is solved by the Minimal Supersymmetric Standard Model (MSSM) and gauge coupling unification is achieved, models with an elementary Higgs particle remain unsatisfactory, because they not only provide no insight into a theory of flavor [12], they also screen any theory of flavor so effectively that experiments at the GUT scale would be required to shed any light on the open questions of the physics of flavor, in particular the ratios of fermion masses and the fermion mixing angles.

A strongly interacting weak sector that generates EWSB dynamically has a greater potential of explaining the physics of flavor. In such a theory it is conceivable that everything can be calculated from (renormalized) Clebsh-Gordan coefficients in a tower of spontaneous symmetry breakings. Such a reduction from the continuous manifold of free parameters to a discrete set of symmetry breaking patterns allowed by group theory would increase the predictive power of theoretical elementary particle physics immensely.

Today we have no working theory of flavor that could explain the ratios of fermion masses and the values of the mixing angles. We do not know if

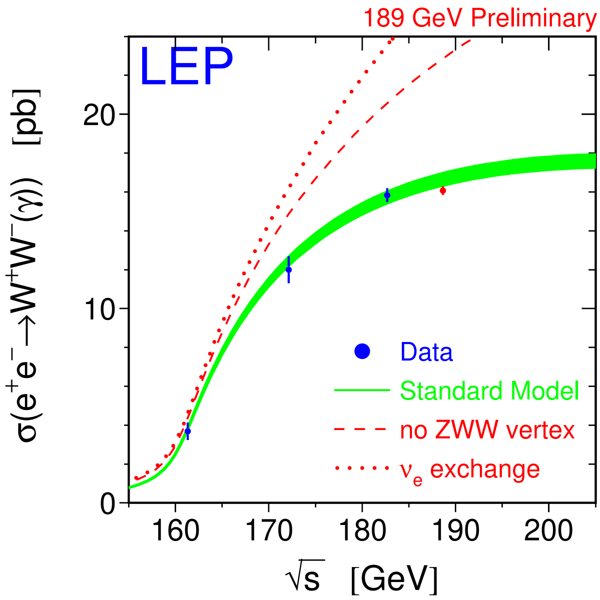


Figure 1.3: *W^+W^--production cross section [6] close to the threshold is in spectacular agreement with the SM prediction. Already at* 183 GeV*, the need for a $W^+W^-Z^0$ vertex was clearly visible.*

Nature was kind enough to hide the answer at the TeV-scale. Of course, she is under no obligation to do so and could as well have used supersymmetry to put the answers out of our experimental reach somewhere between the GUT and Planck scales. The remarkable success of the MSSM with no intermediate mass scales and gauge coupling unification [11] might indeed have been the first direct evidence of such a grand desert.

1.3 The Mission of the Linear Collider

From the previous section, we can identify the main missions for a future LC. Assuming that neither LEP2, TEV33, or LHC have done so, the LC has to try to *discover* the Higgs particle and supersymmetric partners of the particles in the SM. If a Higgs particle has already been found at the time of commissioning, the LC will be in a unique position to determine its nature

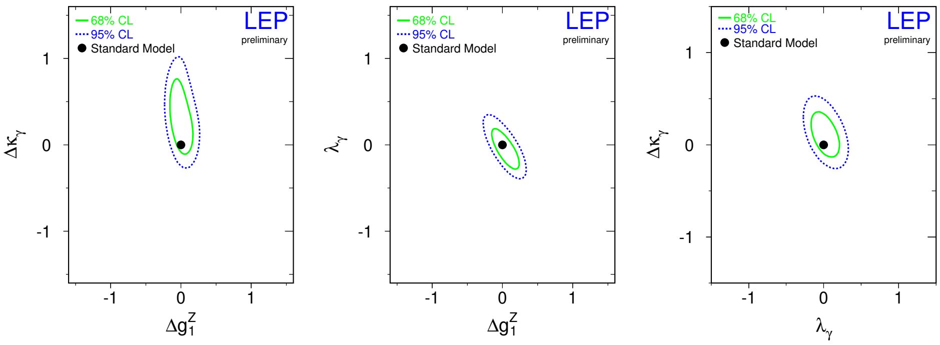


Figure 1.4: *Combined LEP2 and Tevatron limits on Triple Gauge Couplings at the time of the ICHEP98 conference in Vancouver, B.C. (July 1998) [13]. The SM values are favored, but the* 90% *CL boundaries are still* $O(1)$. *See figure 1.5 for a dramatic improvement (note the different scale!).*

by measuring its properties, such as branching ratios, that are difficult to access at a hadron collider, in detail.

Complementary to this discovery mission, the LC has to perform *precision* measurements of the properties of the top quark and of electroweak gauge bosons. Whatever the mechanism of mass generation for the fermions will turn out to be, the top quark couples most strongly of all known fermions to the symmetry breaking sector. In fact, in the minimal SM, the Yukawa coupling of the top quark to the Higgs particle is of the order one. Thus the couplings of the top quark have to be measured as precisely as possible, because they offer the best chance of learning something about the interaction of the symmetry breaking sector with the fermionic sector.

Finally, the LC is in a unique position for precision measurements of the properties of the electroweak gauge bosons. LEP2 has firmly established the trilinear $W^+W^-Z^0$ interaction of gauge bosons. The shape of the W^+W^--production cross section at the threshold as shown in figure 1.3 is as convincing as direct fits of the coupling parameters in figure 1.5.

Due to limited statistics and energy, LEP2 will not be able to measure the $W^+W^-Z^0$ couplings at a level where deviations from the minimal SM can reasonably be expected [14, 15] (see chapter 2). In a strongly interacting model, with no observed light resonances, deviations from the couplings predicted by low energy symmetries are on the order of (see sections 2.7

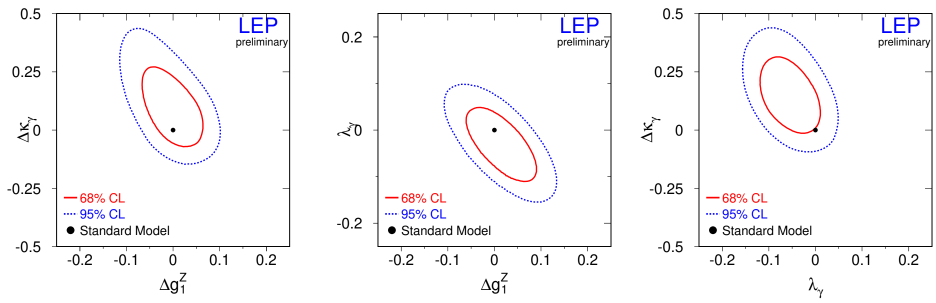


Figure 1.5: *Combined LEP2 and Tevatron limits on Triple Gauge Couplings at the time of the MORIOND 1999 conference (March 1999) [6]. The* 189 GeV *data gives a dramatic improvement over figure 1.4 (note the different scale!) and vanishing TGC are now definitely excluded.*

and A.3.5)

$$\frac{1}{16\pi^2} \approx 6.3 \cdot 10^{-3} \tag{1.2}$$

and in a weakly interacting model, this factor is further reduced by coupling constants. Unless Quantum Field Theory (QFT) breaks down[1] just above the LEP2 energy, the LEP2 experiments are not able to establish any deviation from the SM [18]. The LEP2 measurements are nevertheless important, but we should be disappointed by the "failure" to see any deviation from the SM. The improvements in the determination of the mass of the $W^\pm$-bosons will be more significant, because M_W is not constrained by a low energy symmetry. An improved measurement probes the custodial symmetry, if combined with precision calculations and measurements of M_Z and $\sin^2\theta_w$.

The situation will change qualitatively with the commissioning of the LC. The LC will be able to detect deviations at the level of (1.2) and will be the only machine to do so [1, 2, 3]. Consequently, precision measurements of the properties of electroweak gauge bosons are the final piece of a *no-lose theorem* for a high luminosity ($\mathcal{L} \gtrsim 10^{34}\,\mathrm{cm}^{-2}\mathrm{sec}^{-1}$), high energy ($\sqrt{s} \approx 1\,\mathrm{TeV}$) LC:

Either an elementary Higgs particle and possible supersymmetry is found at the LC as a signal of a weakly interacting symmetry

[1]Recently, the idea of solving the hierarchy problem by introducing large extra dimensions at the TeV-scale [16] has attracted a lot of attention. Such a scenario allows many spectacular signatures that avoid the bounds from Renormalization Group (RG) arguments in four-dimensional QFT [17] and will not be discussed here.

breaking sector or the couplings of electroweak gauge bosons will deviate from the minimal SM predictions at a measurable level in order to accommodate a strongly interacting symmetry breaking sector.

In order to cover the second part of this *no-lose theorem*, the craft of precision electroweak gauge boson physics has to be mastered.

1.4 Roadmap

Antigonus: *Thou art perfect then,*
our ship hath touch'd upon
The deserts of Bohemia?
Mariner: *Ay, my lord, ...*
[William Shakespeare: *The Winter's Tale, III.3*]

Chapter 2 sketches the phenomenological bottom-up approach to the field theoretical description of particle physics experiments and applies it to the SM of electroweak interactions. Most of this material can be found as part of systematic expositions in modern textbooks [19, 20, 21], but is included anyway in a condensed and slightly provocative form in order to put the other chapters in perspective. Chapter 2 attempts to give a concise description that separates the features of the SM of particle physics that are firmly established by phenomenology and the general principles of QFT, on one side, from model dependent assumptions that have not been tested so far, on the other side.

Chapter 2 in particular assumes familiarity with the Renormalization Group (RG) for QFT in the Wilsonian approach [22], roughly on the level of [21]. Therefore, appendix A reproduces a short and deliberately elementary introduction to the Wilson RG for QFT, based on the first half of lectures presented at the German Autumn School for High Energy Physics, Maria Laach, September 9–19, 1997. The physical intuition provided by the Wilson RG is crucial for understanding why the SM is so successful at LEP1 and LEP2 and why we can nevertheless expect a lot of exciting physics at a LC.

Chapters 3 and 4 dive deeper into the more technical parts of the phenomenologists' toolbox. Chapter 3 discusses specific issues arising in the calculation of matrix elements and differential cross sections for gauge bosons in Perturbation Theory (PT). Chapter 4 describes methods for phase space integration and event generation. These chapters combine published original contributions with results that appear to be known only to the practitioners of the field as “tricks of the trade.”

Finally, chapter 5 discusses some aspects of the forthcoming physics of electroweak gauge bosons at a LC.

—2—
The Standard Model as Effective Quantum Field Theory

And as the smart ship grew
In stature, grace, and hue,
In shadowy silent distance grew the Iceberg too.
[Thomas Hardy: *The Convergence of the Twain*]

This chapter explores some consequences of the general principles of QFT for the phenomenology of elementary particles. This material is hardly original (see the texts [19, 20, 21, 23]), but is included in a condensed and intentionally provocative form in order to set the stage for the following chapters. The purpose of this exercise is to allow a separation of kinematics and symmetry on one side, from dynamics on the other side. The result of the application of this formalism to the SM is Janus-faced: many predictions are very solid, because they depend on kinematics and symmetry only, but, by the same token, the independence from dynamical details reveals how little we still know about the underlying dynamics of the SM. This view on the virtues of the SM is purposefully pessimistic. It demonstrates that, despite the tremendous achievements of the past, there is still a lot of exciting physics waiting to be discovered at the high energy frontier.

Perturbative renormalizability has often been presented as the main guiding principle behind the construction of the SM. But when we acknowledge our ignorance about the "true" short distance physics that is out of the reach of the currently operative, as well as the currently planned, colliders, we can not adopt it as a guiding principle anymore. After all, QFT might just be cut off at some string scale. However, as stressed in section A.2.5, among all equivalent theories that describe Nature, there is *always* one perturbatively renormalizable QFT, that contains only relevant and marginal interactions at the highest energy scale. Since perturbatively renormalizable QFTs are a technically more convenient starting point, it is often more economical to

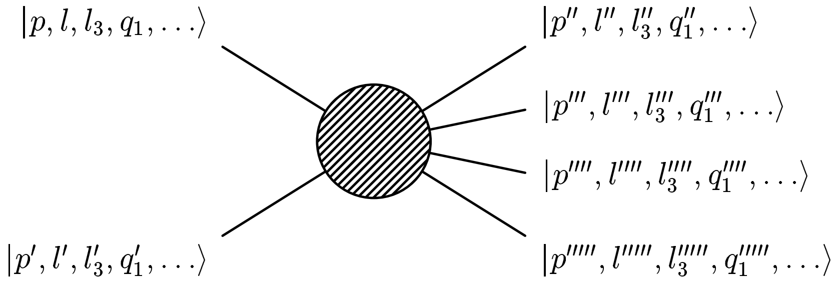


Figure 2.1: *Scattering amplitude.*

use them as the initial condition. Nevertheless, irrelevant interactions will be observed at low energies and one has to explain them as the result of matching conditions at intermediate thresholds as described in section A.3.3 in order to show the consistency of the description.

A comprehensive and systematical modern treatment of all the theoretical issues is provided by Weinberg's text [19, 20]. A more pedagogical treatment from a similarly modern perspective is given in [21], while more on the phenomenology can be found in [24].

2.1 Currents

current: *in general circulation or use; ...*
[Concise Oxford Dictionary]

The *asymptotic single particle states* in a scattering experiment are characterized by quantum numbers belonging to the Poincaré group (energy $E = P_0$, momentum $P_{1,2,3}$, and angular momentum L_3, $\vec{L}^2$) as well as by internal quantum numbers corresponding to conserved charges Q_n:

$$P_\mu \left|p, l, l_3, q_1, \ldots, q_n\right\rangle = p_\mu \left|p, l, l_3, q_1, \ldots, q_n\right\rangle \tag{2.1a}$$

$$\vec{L}^2 \left|p, l, l_3, q_1, \ldots, q_n\right\rangle = l(l+1) \left|p, l, l_3, q_1, \ldots, q_n\right\rangle \tag{2.1b}$$

$$L_3 \left|p, l, l_3, q_1, \ldots, q_n\right\rangle = l_3 \left|p, l, l_3, q_1, \ldots, q_n\right\rangle \tag{2.1c}$$

$$Q_n \left|p, l, l_3, q_1, \ldots, q_n\right\rangle = q_n \left|p, l, l_3, q_1, \ldots, q_n\right\rangle . \tag{2.1d}$$

For this to make sense, the Q_n have to commute with all generators of the Poincaré group[1]

$$[P_\mu, Q_n] = 0 \tag{2.2a}$$

[1] We do not include supersymmetry charges among the Q_n.

$$[L_{\mu\nu}, Q_n] = 0 \tag{2.2b}$$

and in particular

$$[H, Q_n] = 0\,. \tag{2.2c}$$

In general, the Q_n will represent a compact Lie algebra

$$[Q_{n_1}, Q_{n_2}] = \sum_{n_3} i f^{Q}_{n_1 n_2 n_3} Q_{n_3} \tag{2.3}$$

that factorizes into simple non-abelian and abelian subalgebras. The quantum numbers (2.1) define the asymptotic states in scattering experiments completely, if we employ the cluster decomposition principle for asymptotic multi particle states. However, the operators appearing in (2.1) are not sufficient for a Lorentz invariant description of the *interaction* of the states corresponding to tensor products of asymptotic single particle states.

The creation and annihilation operators in the Fock space built from the states in (2.1) are non-local, i. e. their Fourier transforms neither commute nor anti-commute at space like distances. However, the only known way (see [19] for a discussion) of constructing a relativistic (i. e. Lorentz invariant) scattering matrix uses a *local* interaction, i. e. a Lagrangian density that is a Lorentz invariant polynomial of *local* operators that commute or anti-commute at space like distances

$$[\phi(x), \phi'(x')]_{\pm} = 0 \quad \text{for } (x - x')^2 < 0 \tag{2.4}$$

and their derivatives. Therefore, the particle creation and the anti-particle annihilation operators have to be combined into local *field operators* that satisfy (2.4). As a result, it is impossible to have interactions that conserve particle number and the correct description of relativistic scattering will involve an infinite number of degrees of freedom.

The most profound physical consequence of the mathematical result that there are an infinite number of degrees of freedom is the phenomenon of Spontaneous Symmetry Breaking (SSB). For a finite number of degrees of freedom, Wigner has shown long ago that all symmetries of the interaction have to be symmetries of the ground state and the symmetry has to be represented linearly. However, Wigner's theorem breaks down for an infinite number of degrees of freedom and symmetries can be broken spontaneously, guaranteeing the existence of massless scalar particles through Goldstone's theorem. This allows non-linear realizations of these symmetries [25]. The Goldstone bosons are protected from obtaining large masses through radiative corrections, unlike regular scalars. On the other hand, non-linearly realized gauge symmetries provide a consistent description of massive spin-1 particles that does not suffer from bad high-energy behavior (see section 2.6 for a detailed discussion).

The charges Q_n can be constructed from local quantities as the integral of a charge density $j_n^0(x)$ on a space-like hypersurface $x_0 = t$

$$Q_n(t) = \int\limits_{x_0=t} \mathrm{d}^3x\, j_n^0(x)\,, \tag{2.5}$$

where $j_n^0(x)$ is the time component of a conserved current $j_n^\mu(x)$

$$\partial_\mu j_n^\mu(x) = 0\,. \tag{2.6}$$

If the matrix elements of the $j_n^\mu(x)$ fall off sufficiently rapidly at large distances, the charge Q_n is conserved

$$\frac{\mathrm{d}Q_n}{\mathrm{d}t}(t) = \int\limits_{x_0=t} \mathrm{d}^3x\, \partial_0 j_n^0(x) = \int\limits_{x_0=t} \mathrm{d}^3x\, \vec\nabla \vec{j}_n(x) = 0\,. \tag{2.7}$$

These conserved currents are particularly useful in the phenomenology of interacting field theories, because they are not renormalized. For the non abelian factors of the Lie algebra generated by the charges Q_n, this is trivial, because any renormalization $j_\mu \to Zj_\mu$ would induce a renormalization $Q \to ZQ$ and violate the commutation relations (2.3). For the abelian factors, we have to make the further assumption that there are operators ψ that form a non trivial representation

$$[Q_n, \psi] = q_{n,\psi}\psi\,. \tag{2.8}$$

In this relation, any renormalization of the ψ will cancel and we find again a non renormalization theorem for the current.

Matrix elements of the conserved currents in asymptotic states are constrained by symmetry, independent of dynamical assumptions. For example, the matrix element of a current between two scalar single particle states with identical mass ($p^2 = p'^2$) must have the form[2]

$$\langle p', q'_n | j_{n,\mu}(0) | p, q_n \rangle = (p'_\mu + p_\mu) q_n F((p'-p)^2)\delta_{q'_n q_n}\,, \quad F(0) = 1 \tag{2.9}$$

from Lorentz symmetry and current conservation alone, where the normalization of the form factor at zero momentum transfer $F(0) = 1$ is fixed by

[2]The eigenstates are orthogonal for hermitian charges Q: $\langle q'|Q|q\rangle \propto \delta_{q,q'}$.

$$
\begin{aligned}
q_n 2p_0 (2\pi)^3 \delta^3(\vec{p'} - \vec{p}) &= q_n \langle p', q_n | p, q_n \rangle = \langle p', q_n | Q_n | p, q_n \rangle \\
&= \int\limits_{x_0=t} \mathrm{d}^3 x \; \langle p', q_n | j_{n,0}(x) | p, q_n \rangle = \int\limits_{x_0=t} \mathrm{d}^3 x \; e^{i\vec{x}(\vec{p}-\vec{p'})} \langle p', q_n | j_{n,0}(0) | p, q_n \rangle \\
&\qquad\qquad\qquad\qquad = (2\pi)^3 \delta^3(\vec{p'} - \vec{p}) \langle p', q_n | j_{n,0}(0) | p, q_n \rangle \; ,
\end{aligned}
$$

i. e.

$$
q_n 2p_0 = \langle p, q_n | j_{n,0}(0) | p, q_n \rangle \; . \tag{2.10}
$$

Similar relations can be derived for matrix elements of particles with spin, but in general there is a larger number of independent form factors for which symmetry provides less constraints.

2.2 Weak Interactions

The low energy manifestations of the interactions described by the Standard Model (SM) are the long range electromagnetic interaction and the short range weak interaction. In Nature, we observe weak interactions in the burning of the solar fuel

$$
4p + 2e^- \to {}^4\mathrm{He} + 2\nu_e + 26.73\,\mathrm{MeV} - E_\nu \, . \tag{2.11}
$$

While the released energy is the binding energy of the ^{4}He nucleus due to the strong interaction of protons and neutrons, the process (2.11) would be impossible without the weak interaction, because the strong interaction conserves isospin, which forbids $p+e^- \to n+\nu_e$ transitions. The other natural manifestation of weak interactions is in the nuclear β-decay, for example

$$
{}^{137}\mathrm{Cs} \to {}^{137}\mathrm{Ba} + e^- + \bar{\nu}_e \, . \tag{2.12}
$$

Decades of particle physics experiments with accelerators have found many more examples of weak interactions that can be observed because they violate the isospin and parity invariance of strong interactions (see [9] for the most exhaustive and up-to-date list).

2.3 Fermi Theory

αὐτός ἔφα
[Pythagoras' pupils]

As mentioned in the previous section, there are two defining characteristics of weak interactions at low energies: parity violation and isospin non-conservation. The former is the result of the experimental observation that all electrons (and muons and taus) produced in β-decay are left handed, while all positrons (and anti-muons and anti-taus) produced in β^+-decay are right handed[3]. Therefore, the weak interaction couples in the leptonic sector to $(V - A)$ charged currents of the form

$$\bar{\nu}\gamma_\mu(1-\gamma_5)e + \text{h.c.}\,.$$

Using the $\tau^\pm = (\tau_1 \pm i\tau_2)/2$ isospin shift operators, such charged currents can be combined in the $\pm$-components of a conserved isospin vector current

$$\begin{aligned} j^\pm_\mu(x) = (\bar{\nu}(x)\ \bar{e}(x)) & \left[\tau^\pm \otimes \gamma_\mu \frac{1-\gamma_5}{2}\right] \begin{pmatrix}\nu(x)\\ e(x)\end{pmatrix} \\ & + (\bar{\nu}_\mu(x)\ \bar{\mu}(x)) \left[\tau^\pm \otimes \gamma_\mu \frac{1-\gamma_5}{2}\right] \begin{pmatrix}\nu_\mu(x)\\ \mu(x)\end{pmatrix} \\ & \quad + (\bar{\nu}_\tau(x)\ \bar{\tau}(x)) \left[\tau^\pm \otimes \gamma_\mu \frac{1-\gamma_5}{2}\right] \begin{pmatrix}\nu_\tau(x)\\ \tau(x)\end{pmatrix}\,, \end{aligned} \tag{2.13}$$

which is written more compactly as

$$\vec{j}_\mu(x) = \bar{\psi}\left[\frac{\vec{\tau}}{2} \otimes \gamma_\mu \frac{1-\gamma_5}{2}\right]\psi\,, \tag{2.14}$$

where the summation over all isospin doublets is implied. The neutral component $j^0_\mu(x)$ of the current (2.14) is masked at low energies by much stronger electromagnetic interactions of charged leptons and strong interactions of hadrons and was observed only much later, after neutrino beams became available.

Making the educated guess that the hadronic part is described by a similar isospin current, the simplest short range interaction imaginable couples these currents

$$L(x) = \frac{g}{\Lambda^2} j^+_\mu(x) j^{-,\mu}(x) \tag{2.15}$$

with g a suitable coupling constant and Λ a suitable scale. According to section A.2.5, the interaction (2.15) is irrelevant and expected to be very weak at low energies (compared to Λ).

[3]This observation summarizes many complicated experiments and ingenious theoretical interpretations.

The interaction (2.15) has been tested quantitatively (including radiative corrections, see e. g. [24]) in the leptonic sector, where the matrix elements of the products of isospin currents can be calculated without further assumptions. From the comparisons with experimental results, lepton universality, i. e. that each doublet of leptons contributes *identically* to the current as in (2.13), is firmly established.

For the hadronic part of the current, no ab-initio calculations are possible yet, due to our limited understanding of the non-perturbative dynamics of the strong interactions. In semi-leptonic processes like $B^+ \to \bar{D}^0\mu^+\nu_\mu$, the matrix element of the interaction can be factorized in terms of a leptonic and a hadronic current

$$\left\langle \mu^+\nu_\mu\bar{D}^0 \middle| j^+_\mu(x) j^{-,\mu}(x) \middle| B^+ \right\rangle = \left\langle \mu^+\nu_\mu \middle| j^+_\mu(x) \middle| 0 \right\rangle \left\langle \bar{D}^0 \middle| j^{-,\mu}(x) \middle| B^+ \right\rangle . \quad (2.16)$$

Model independent symmetry arguments for the conserved vector current and partially conserved axial current similar to (2.9) can be used to express (2.16) in terms of a single unknown form factor. For heavy hadrons, the form factor has a known normalization a zero momentum transfer, up to small corrections [26]. This provides non-trivial cross checks for such processes and the interaction (2.15) has passed these tests. Similar symmetry arguments have less predictive power for light baryons and vector mesons, because there are more independent form factors, but some non-trivial checks can still be performed.

Compared to this, purely hadronic weak processes remain poorly understood, because the matrix elements can not be factorized in terms of hadronic currents and ab-initio calculations in terms of quarks are not yet possible. However, this reflects more on our still poor understanding of non-perturbative QFT than on the accuracy of the description of weak interactions with the Fermi interaction (2.15).

2.4 Flavor Symmetries

What are the hadronic currents that the weak interaction couples to? A comprehensive study of the transitions among hadrons induced by the weak interaction currents

$$j^a_{\mu,\text{meson}} = \sum_{j,k} i\Phi^\dagger_j T^a_{jk} \overleftrightarrow{\partial_\mu} \Phi_k \quad (2.17a)$$

$$j^a_{\mu,\text{baryon}} = \sum_{j,k} \bar{\Psi}_j T^a_{jk} \Gamma_\mu \Psi_k \quad (2.17b)$$

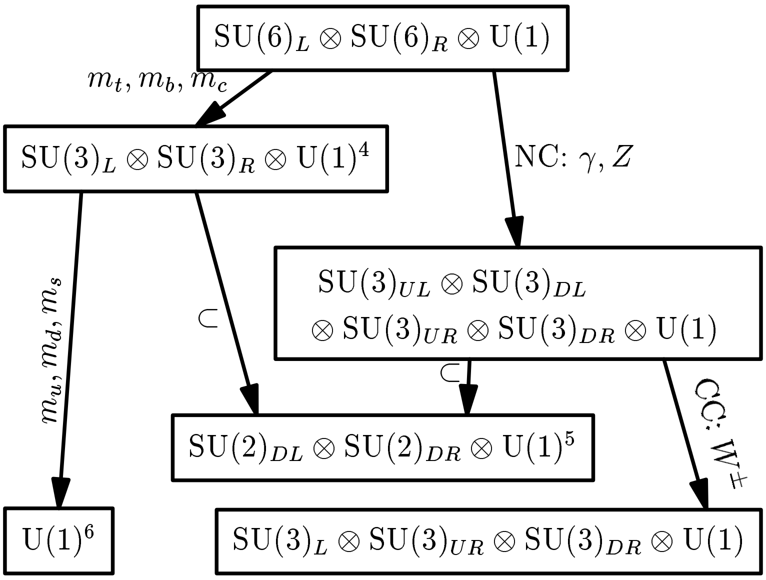


Figure 2.2: *Chiral symmetry breaking patterns of hadrons by weak interactions and quark masses. On the left hand side, the symmetry is broken by an as yet unknown mechanism that creates heavy quark masses. In the right hand side, the embedding of the weak interaction gauge group in the global chiral symmetry groups results in quite different symmetry breaking patterns.*

together with the hadron masses M_j

$$L_{0,\text{meson}} = \sum_j \left[\frac{1}{2} (\partial_\mu \Phi_j)^\dagger (\partial^\mu \Phi_j) - \frac{1}{2} M_j^2 \Phi_j^\dagger \Phi_j \right] \tag{2.18a}$$

$$L_{0,\text{baryon}} = \sum_j \left[i\bar{\Psi}_j \partial\!\!\!/ \Psi_j - M_j \bar{\Psi}_j \Psi_j \right] \tag{2.18b}$$

reveals intriguing symmetry patterns for the meson fields Φ_j and baryon fields Ψ_j, which can be summarized by the succession of global (i. e. ungauged) SU(N) flavor symmetry groups in figure 2.2[4]. The strong interac-

[4]Note that the $SU(6)_{L,R}$ symmetry groups in figure 2.2 contain some revisionistic historiography, because no hadrons with top flavor are observed in nature. Furthermore, the Fermi interaction is not applicable at energies as high as the top threshold. Therefore, a strictly phenomenological analysis would use $SU(5)_{L,R}$ instead. Nevertheless, for the purpose of studying the symmetry properties, it is useful to pretend temporarily that the top is lighter than it actually is.

tions appear to be completely flavor blind and hadrons can be classified in representations of a global $\mathrm{SU}(6)_L \otimes \mathrm{SU}(6)_R \otimes \mathrm{U}(1)$ symmetry group[5].

The classification of the hadronic currents and their interactions according to the symmetries described in figure 2.2 is complicated by the fact that the observed hadrons do not live in the fundamental representation N of $\mathrm{SU}(N)$. Instead they live in the irreducible components of product representations that carry no overall charge of the color SU(3), responsible for Quantum Chromodynamics (QCD), the strong interaction in the quark picture. These representations of $\mathrm{SU}(N)$ are $N \otimes \bar{N} = 1 \oplus (N^2 - 1)$ for the mesons and $N \otimes N \otimes N$ for the baryons. However, the leptons do form a

[5]The six different flavors will later correspond to quarks, coming with left handed and right handed chiralities. Note that the SU(6)'s must not be confused with the nonrelativistic SU(6) symmetry containing the SU(3) of the light hadrons and the SU(2) of spin angular momentum.

Table 2.1: *The Standard Model (SM) of elementary particle physics. We have color ($\mathbf{C}$) singlets ($\mathbf{1}$) and triplets ($\mathbf{3}$), weak isospin ($\mathbf{T}$) singlets and doublets ($\mathbf{2}$) and various hypercharge (Y) assignments.*

	Leptons		$(\mathbf{C},\mathbf{T})_Y$	$Q = T_3 + \frac{Y}{2}$
$\nu_{e,R}\,(?)$	$\nu_{\mu,R}\,(?)$	$\nu_{\tau,R}\,(?)$	$(\mathbf{1},\mathbf{1})_0$	0
e_R	μ_R	τ_R	$(\mathbf{1},\mathbf{1})_{-2}$	-1
$\begin{pmatrix}\nu_{e,L}\\ e_L\end{pmatrix}$	$\begin{pmatrix}\nu_{\mu,L}\\ \mu_L\end{pmatrix}$	$\begin{pmatrix}[\nu_{\tau,L}]\\ \tau_L\end{pmatrix}$	$(\mathbf{1},\mathbf{2})_{-1}$	$\begin{pmatrix}0\\ -1\end{pmatrix}$
	Quarks			
u_R	c_R	t_R	$(\mathbf{3},\mathbf{1})_{4/3}$	$\frac{2}{3}$
d_R	s_R	b_R	$(\mathbf{3},\mathbf{1})_{-2/3}$	$-\frac{1}{3}$
$\begin{pmatrix}u_L\\ d_L\end{pmatrix}$	$\begin{pmatrix}c_L\\ s_L\end{pmatrix}$	$\begin{pmatrix}t_L\\ b_L\end{pmatrix}$	$(\mathbf{3},\mathbf{2})_{1/3}$	$\begin{pmatrix}\frac{2}{3}\\ -\frac{1}{3}\end{pmatrix}$
	Gauge Bosons			
γ Z^0	$W^\pm$	g		
	Higgs			
	$\Phi\,(?)$		$(\mathbf{1},?)_?$	$?$

fundamental representation of the symmetries in figure 2.2, as shown in the upper left part of table 2.1. Furthermore, there is independent experimental evidence, in particular from deep-inelastic lepton hadron scattering, for quarks as constituents of hadrons, which live in a fundamental representation (see table 2.1). In particular, the properties of hadrons with heavy flavors are well described by heavy quarks surrounded by a $\mathrm{SU}(3)_L \otimes \mathrm{SU}(3)_R$ flavor symmetric "cloud" of light quarks and gluons [26]. The product representations for the observed hadron spectra are therefore understood as the bound states of quarks.

The $\mathrm{SU}(6)_L \otimes \mathrm{SU}(6)_R$ symmetry is badly broken by the heavy flavor masses m_t, m_b, and m_c, which leaves the $\mathrm{SU}(3)_L \otimes \mathrm{SU}(3)_R$ of the eightfold way of the three light flavors [27]. This is already a useful symmetry that can be used to derive quantitative results, provided that the breaking of the flavor-SU(3) to the isospin-SU(2) due to the finite strange quark mass is taken into account properly, together with the spontaneous chiral symmetry breaking $\mathrm{SU}(3)_L \otimes \mathrm{SU}(3)_R \to \mathrm{SU}(3)_{L+R}$ caused by QCD. The product representations are the familiar singlet plus octet for the light mesons $3 \otimes \bar{3} = 1 \oplus 8$ and the singlet, two octets and a decuplet for the light baryons $3 \otimes 3 \otimes 3 = 1 \oplus 8 \oplus 8 \oplus 10$.

On the other hand, the $\mathrm{SU}(6)_L \otimes \mathrm{SU}(6)_R$ is also broken by differing charges for weakly interacting neutral current interactions (i. e. Z^0-exchange and QED) for up-type quarks and down-type quarks. The remaining $\mathrm{SU}(3)_{UL} \otimes \mathrm{SU}(3)_{UR}$ corresponds to the flavors u, c, and t of the up-type quarks and the $\mathrm{SU}(3)_{DL} \otimes \mathrm{SU}(3)_{DR}$ corresponds to the flavors d, s, and b of the down-type quarks. This symmetry in broken further by the weakly interacting charged current interactions to a $\mathrm{SU}(3)_L$ for the left handed part that interacts and a $\mathrm{SU}(3)_{UR} \otimes \mathrm{SU}(3)_{DR}$ for the non-interacting right handed parts. Again, these symmetries are badly broken by heavy flavor masses, but the $\mathrm{SU}(3)_L \otimes \mathrm{SU}(3)_R$ and the $\mathrm{SU}(3)_{UL} \otimes \mathrm{SU}(3)_{DL} \otimes \mathrm{SU}(3)_{UR} \otimes \mathrm{SU}(3)_{DR}$ have a common subgroup $\mathrm{SU}(2)_{DL} \otimes \mathrm{SU}(2)_{DR}$ that is a good symmetry and has interesting phenomenological consequences [28].

An additional complication is provided by the fact that the $\mathrm{SU}(3)_{UL} \otimes \mathrm{SU}(3)_{UR}$, the $\mathrm{SU}(3)_{DL} \otimes \mathrm{SU}(3)_{DR}$, and the isospin SU(2) of the weak interactions twisted by the CKM mixing matrix. Without the intuition provided by quarks, this low energy symmetry structure of the weak interactions would be very hard to understand.

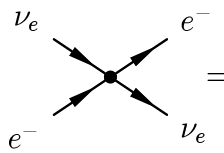
The SM provides a complete phenomenological description of the currents described in this section. However, of all the symmetry breakings in figure 2.2, it only explains the two steps on the right, which are due to neutral currents and charged currents. The other symmetry breakings, as well as the CKM mixing have to be put in by hand, introducing most of the free parameters in the SM.

The strong interaction physics that is responsible for the relation between the fundamental representation and the product representations is interesting in itself, but not of fundamental importance for the weak interaction physics under consideration. The significant observation of the present section is that the leptonic and semi-leptonic low energy weak interactions can indeed be described by the Fermi interaction (2.15) for observable hadronic and leptonic currents. The axial part of the short range neutral current weak interaction has also been observed more recently in low energy experiments through their parity violating effects. Advances in experimental atomic physics have made atomic parity violation a viable tool for precision physics [29].

The emerging picture of the weak interactions at low energy is that of interacting isospin SU(2) currents, where the charged components act on the left handed fields and the neutral component couples differently to the two members of the isospin doublet particles. The symmetry is incomplete, as long as the charged current and neutral current pieces of the interaction appear to be unrelated. Below in section 2.7, we will see that the corresponding gauge bosons are indeed related by a custodial $\mathrm{SU}(2)_c$ symmetry [30].

2.5 Higher Energy Scattering

Having established the Fermi interaction (2.15) for low energy weak interactions, we must study if and when this description will break down at higher energies, either phenomenologically or from theoretical inconsistencies. From power counting, the interaction is irrelevant and the RG flow in the ultraviolet will eventually leave the perturbative domain (see section A.2.4). But instead of studying radiative corrections, we can also study a tree level process like $e\nu$ scattering. The differential cross section is found to be quadratically rising with energy



$$\Longrightarrow \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{G_F^2}{4\pi^2} \cdot E_{CM}^2 \,. \tag{2.19}$$

On the other hand, we can project the scattering amplitude on the S-wave component and verify the unitarity of the partial wave

$$|\mathcal{M}_{J=0}| \leq 1 \,. \tag{2.20}$$

This gives an upper bound on differential cross section

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \leq \frac{1}{E_{CM}^2} \,. \tag{2.21}$$

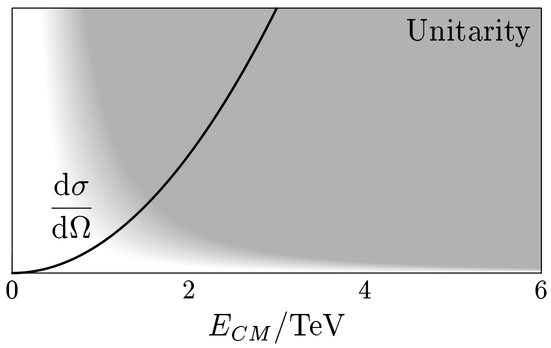


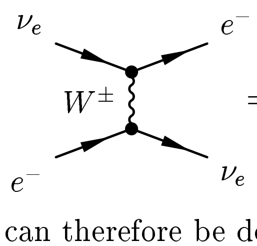
Figure 2.3: *Differential cross section and S-wave tree-level unitarity bound for* $e^-\nu_e \to e^-\nu_e$.

Both the differential cross section and the unitarity bound are shown in figure 2.3, where the unitarity bound is turned on smoothly, because we have to allow for higher order corrections close to the bound. In any case, it is obvious from figure 2.3, that the Fermi interaction (2.15) has to be replaced before 1 TeV.

The solution to this problem is provided by softening the interaction at high energies, e. g.

$$\frac{1}{\Lambda^2} \to \frac{-1}{p^2 - \Lambda^2} \tag{2.22}$$

for some combination p^2 of kinematical invariants that scales as E_{CM}^2. The particular example in (2.22) is just one of many possibilities that reduce to the Fermi interaction at low energy. However, it corresponds to replacing the local interaction (2.15) by the exchange of vector bosons



$$\Longrightarrow \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{g^4}{32\pi^2}\frac{1}{E_{CM}^2}\frac{1}{\left(1 - \cos\theta + \frac{2M_W^2}{E_{CM}^2}\right)^2} \tag{2.23}$$

and can therefore be described again by a local interaction. Unlike an arbitrary form factor, a local interaction produces a Lorentz invariant scattering matrix (see section 2.1). The formal Lorentz invariance of an ad-hoc form factor in a non-local interaction does *not* suffice for Lorentz invariance of the scattering matrix for all matrix elements if the form factor is taken into account consistently [19]. Space-like correlations will ruin Lorentz invariance.

Incidentally, the interaction that produces the vector boson exchange to replace the irrelevant current-current interaction,

$$L(x) = \sum_a A_\mu^a(x) j^{a,\mu}(x), \tag{2.24}$$

is marginal (see A.2.5) and the scale in the irrelevant interaction is replaced by the mass of the new particle. This fits of course nicely within the EFT paradigm, where a low energy EFT breaks down at the scale where new particles and interactions have to be introduced.

2.6 Intermediate Vector Bosons

The introduction of intermediate vector bosons (2.23) is not without problems, because unitarity requires that all particles that are exchanged must also appear in external states. Massive vector bosons have three degrees of freedom and massless vector bosons (photons) have only two degrees of freedom. However, a naive counting of the components of a Lorentz vector finds four degrees of freedom.

One approach to a consistent theory would construct the vector bosons as narrow spin one resonances of scalars or spin-1/2 fermions. In such a theory, unitarity could be proven in a Hilbert space of asymptotic states built from scalars and spin-1/2 fermions, avoiding all problems with unphysical degrees of freedom. However, constituents that could form resonances with the correct quantum numbers are not observed and Occam's razor suggests to attempt a description in terms of the vectorial degrees of freedom before inventing new constituents.

The current operators have all the right quantum numbers, but they have a mass dimension of three and can therefore not be identified with intermediate vector bosons with a mass dimension of one. Unfortunately, the very fact that conserved currents are not renormalized prevents them from acquiring an anomalous dimension and turning into a vector field with canonical dimension for sufficiently large coupling[6].

For non-interacting massive vector bosons, a simple solution is that the equation of motion

$$(\partial^2 + M^2) A^\mu - \partial^\mu \partial_\nu A^\nu = 0 \tag{2.25}$$

derived from the Proca Lagrangian with an explicit mass term

$$L = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} M^2 A_\mu A^\mu, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.26}$$

[6] See however [31] for an interesting speculation along this line.

eliminates the unphysical degree of freedom automatically, because of

$$M^2 \partial_\mu A^\mu = 0 \,. \tag{2.27}$$

The operator equation (2.27) removes precisely the degree of freedom that would correspond to a negative norm state in a naive canonical quantization of all four components.

2.6.1 Power Counting

If the vector boson A is coupled to a *conserved* current in (2.24), the operator equation (2.27) remains intact. However, even if the vector boson couples to conserved currents only, it is not true in general that contributions from the longitudinal part of the propagator

$$-\frac{i}{p^2 - M^2 + i\epsilon} \left(g_{\mu\nu} - \frac{p_\mu p_\nu}{M^2} \right) \tag{2.28}$$

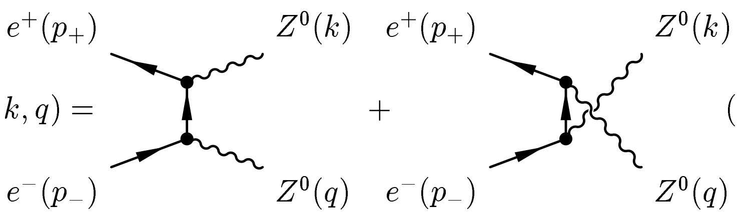
cancel in physical amplitudes. Instead, except for exceptional cases, (2.27) has to be enforced explicitely in intermediate states. As a consequence, the power counting rules (A.89) used to classify effective Lagrangians are violated by the longitudinal polarization state.

If the currents in (2.17) are conserved and correspond to an abelian symmetry, i. e.

$$\left[T^a, T^b \right] = 0 \,, \tag{2.29}$$

the longitudinal polarization state decouples nevertheless. This is trivial from the equations of motion for on-shell currents, but in order to show that the contributions from off-shell propagators cancel, one has to sum over all Feynman diagrams (see also section 3.2). For example, in the process $e^+e^- \to Z^0 Z^0$, the diagrams

$M_{\mu\nu}(p_+, p_-; k, q) =$



(2.30)

contribute equally to the Ward identity

$$k^\mu \epsilon^\nu(q) M_{\mu\nu}(p_+, p_-; k, q) = 0 \,. \tag{2.31}$$

In general, (2.29) is satisfied for the neutral current piece of the weak interaction, because these particular T^a are generated from the Pauli matrices $\mathbf{1}$ and τ_3 by linear combination, direct product, and direct sum.

Due to this decoupling, it is possible to replace the Proca propagator (2.28) for abelian currents by a Stückelberg propagator

$$-\frac{i}{p^2 - M^2 + i\epsilon}\left(g_{\mu\nu} - (1-\xi)\frac{p_\mu p_\nu}{p^2 - \xi M^2 + i\epsilon}\right)\,, \tag{2.32}$$

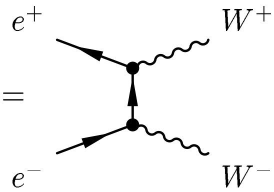
which is compatible with the power counting rules (A.89). Then the choice $\xi = 1$ is particularly convenient for quick calculations, while the cancellation of ξ dependent pieces provides useful cross checks in more complicated calculations. With (2.32), ∂A can be eliminated consistently from matrix elements

$$\langle \text{phys.}|\partial_\mu A^\mu(x)|\text{phys.}'\rangle = 0 \tag{2.33}$$

as a *free* field, but it does not necessarily vanish as an operator in the full indefinite metric Hilbert space.

The situation is more involved for charged current interactions in which (2.29) is violated. For example, in the process $e^+e^- \to W^+W^-$ only one of the diagrams corresponding to (2.30) does not vanish

$$M^{(1)}_{\mu\nu}(p_+, p_-; k, q) = \tag{2.34}$$

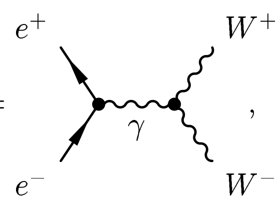


and it does *not* satisfy the Ward identity by itself

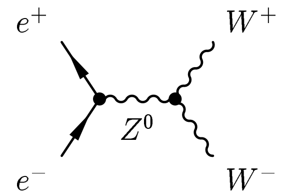
$$k^\mu \epsilon^\nu(q) M^{(1)}_{\mu\nu}(p_+, p_-; k, q) \propto \bar v(p_+)\not\epsilon(q) u(p_-)\,. \tag{2.35}$$

Since the $W^\pm$ are charged, there is a second diagram with photon exchange

$$M^{(2)}_{\mu\nu}(p_+, p_-; k, q) = \quad , \tag{2.36}$$



but this does not cancel (2.35) for minimal coupling $\partial_\mu \to \partial_\mu - ieA_\mu$ of the $W^\pm$. Therefore, we must *not* use a Stückelberg propagator (2.32) for massive vector bosons coupled to charged currents naively. In an attempt to salvage the power counting, we can take a hint from unbroken non-abelian gauge theories (a. k. a. Yang-Mills theories) and investigate another diagram



$$, \tag{2.37}$$

arising from a coupling of the $W^\pm$'s to a third gauge boson, the Z^0, which couples to the τ_3 component of the current, in order to close the gauge algebra generated by the $\tau^\pm$. Indeed using the non-abelian gauge vertex and putting the outgoing $W^\pm$'s on their mass-shell

$$k^\mu \epsilon^\nu(q) M^{(1)}_{\mu\nu}(p_+, p_-; k, q) \propto \frac{s - M_W^2}{s - M_Z^2} \bar v(p_+) \not\epsilon(q) u(p_-) \,. \tag{2.38}$$

Thus (2.35) could be canceled with charge assignments that are compatible with the gauge algebra, as long as $M_W = M_Z$. However, the sum of (2.34), (2.36) and (2.37) *still* couples to unphysical polarization states if the external bosons are massive, because it is impossible to cancel the massless photon pole this way. Moreover, adding the Yang-Mills interaction to the Proca Lagrangian (2.26) results in equations of motion that violate (2.27). Therefore, a successive plugging of the holes is not likely to succeed and we should take a step back and reconsider, how unitarity is implemented in covariantly quantized *massless* Yang-Mills theories.

In addition to the two physical degrees of freedom there are four unphysical degrees of freedom in the covariant quantization of gauge theories: the scalar and longitudinal polarization states $\partial_\mu A^\mu(x)$ and $A_L(x)$ of the gauge bosons are accompanied by two Faddeev-Popov ghosts $\eta(x)$ and $\bar\eta(x)$ [32]. The quartet-mechanism [33] for Becchi-Rouet-Stora (BRS) [34] invariant interactions guarantees that these unphysical degrees of freedom decouple from the two physical degrees of freedom. Only in an abelian gauge theory, as QED, do the unphysical polarization states and ghost decouple separately. Thus the physical Hilbert space in a Yang-Mills theory is not characterized by the simple Gupta-Bleuler condition (2.33), but by the cohomology of the conserved BRS-charge

$$Q_{\mathrm{BRS}}(t) = \frac{1}{2} \int_{x_0=t} \mathrm{d}^3\vec x \, \mathrm{tr} \Bigg(\pi_0(x) \overleftrightarrow{\partial}_0 \eta(x) - ig\pi_0(x) \left[A_0(x), \eta(x)\right] \\ + \frac{ig}{2} \left[\eta(x), \eta(x)\right] \partial_0 \bar\eta(x) \Bigg) \,. \tag{2.39}$$

The cohomology arises because the BRS-charge Q_{BRS} projects on zero norm states

$$\left\|Q_{\mathrm{BRS}}\left|\Psi\right\rangle\right\|^2 = \left\langle\Psi\middle|Q_{\mathrm{BRS}}^2\middle|\Psi\right\rangle = 0\,, \tag{2.40}$$

due to

$$Q_{\mathrm{BRS}}^\dagger = Q_{\mathrm{BRS}} \tag{2.41a}$$
$$Q_{\mathrm{BRS}}^2 = 0\,, \tag{2.41b}$$

and such states have to factored out of the physical states characterized by

$$Q_{\mathrm{BRS}}\left|\mathrm{phys.}\right\rangle = 0\,, \tag{2.42}$$

in order to construct a proper Hilbert space. For massive vector bosons, some other degree of freedom has to take the place of the now physical longitudinal polarization state $A_L(x)$ in the quartet-mechanism. This degree of freedom can be provided by the Goldstone bosons of a spontaneous symmetry breaking and the quantum numbers will match, if, and only if, the broken symmetry is the gauge symmetry of the massive vector bosons.

2.6.2 Hidden Symmetry

Spontaneously broken gauge theories provide a consistent description of massive vector bosons, in which the dimensions of the fields are indeed given by (A.10) and (A.11). There is no argument that they are the *only* consistent description. However, such an argument is not necessary, because it is possible to use a non-linear field redefinition to rewrite any low energy effective interaction of vector bosons as the unitarity gauge version of a non linearly realized gauge theory [35, 36, 37]. In an abelian toy example[7] there are the correspondences

$$\begin{pmatrix}\psi(x)\\ V_\mu(x)\end{pmatrix} \Longleftrightarrow \begin{pmatrix}\chi'(x)\\ \frac{-1}{gf}D_\mu\phi(x)\end{pmatrix} = \begin{pmatrix}e^{-i\phi(x)/f}\chi(x)\\ A_\mu(x)-\frac{1}{gf}\partial_\mu\phi(x)\end{pmatrix}\,. \tag{2.43}$$

χ is a matter field and V_μ is massive vector field, while ϕ is a Goldstone boson field and A_μ a gauge field of a non-linearly realized U(1)

$$\begin{pmatrix}\chi(x)\\ \phi(x)\\ A_\mu(x)\end{pmatrix} \to \begin{pmatrix}e^{i\omega(x)}\chi(x)\\ \phi(x)+f\omega(x)\\ A_\mu(x)+\frac{1}{g}\partial_\mu\omega(x)\end{pmatrix}\,. \tag{2.44}$$

[7] In the non-abelian generalization, the physics is obscured slightly by the need to parametrize the factor group manifold G/H. Nevertheless, it is a straightforward application of the formalism set forth in [25] and explicit formulae are given in [36, 37].

Without a specification of the subsidiary conditions for the quantization of the vector fields, it appears that (A_μ, ϕ) has one more degree of freedom than V_μ. However, V_μ is understood to be quantized with $\partial_\mu V^\mu = 0$ enforced explicitely, while (A_μ, ϕ) is quantized in an indefinite norm Hilbert space where ϕ and $\partial_\mu A^\mu$ conspire with the Faddeev-Popov ghosts to decouple from physical amplitudes [33] and the same number of degrees of freedom survive. Since the non-linear field redefinitions do not change the on-shell scattering matrix elements [38], the two formulations will yield the same predictions for observable quantities.

The preceding reasoning does not prove that all vector bosons are gauge bosons in disguise, but it comes very close, because it shows that for every theory with massive vector bosons, there is another theory in which the vector bosons are parametrized gauge bosons of a non-linearly realized gauge symmetry. These theories are *indistinguishable* at energy scales below the symmetry breaking scale $\Lambda = 4\pi f$ characterizing the non-linear realization. Above the symmetry breaking scale, the two theories are of course very different, because the gauge symmetry will then be realized linearly.

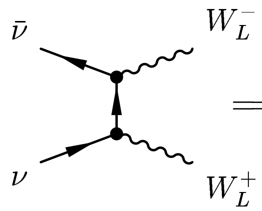
In a covariant R_ξ-gauge, a special case of (2.42) is the new subsidiary condition

$$\langle \text{phys.} | \partial_\mu A^\mu(x) - \xi m_A \phi(x) | \text{phys.}' \rangle = 0\,. \tag{2.45}$$

that takes the place of (2.33). At energies large compared to the boson mass, the longitudinal polarization vector $\epsilon_L^\mu(k)$ can be approximated by k^μ / m_A. Then (2.45) turns into the celebrated *equivalence theorem* [35, 39] and the scattering amplitudes for longitudinal vector bosons can be calculated more efficiently by calculating the corresponding amplitudes for Goldstone bosons.

2.6.3 Tree Level Unitarity

A complementary analysis arrives at the same conclusion. A systematic study of the tree level unitarity of four-point functions and five-point functions calculated with arbitrary interactions reveals that *only* spontaneously broken gauge theories result in amplitudes that satisfy tree level unitarity at high energy [35] (see also [40] for a less systematical but more pedagogical account). As could be expected from the study of the Ward identities, the production of *longitudinal* vector bosons



$$\Longrightarrow \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{G_F^2}{32\pi^2} \cdot E_{CM}^2 \cdot \sin^2\theta \tag{2.46}$$

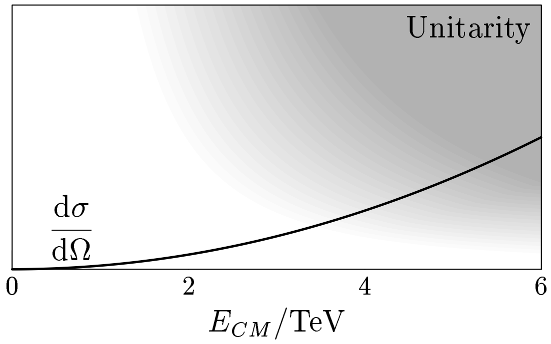


Figure 2.4: *Tree-level unitarity bound for* $\nu\bar{\nu} \to W_L^+ W_L^-$.

is responsible for this result. P-wave unitarity ($|\mathcal{M}_{J=1}| \leq 1$) demands

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \leq \frac{18\sin^2\theta}{E_{CM}^2} \tag{2.47}$$

which is displayed in figure 2.4. Unitarity is maintained up to considerably higher scales than in figure 2.3, but will eventually be violated at a few TeV. Continuing this heuristic “patching up” will unambiguously [35] lead to a spontaneously broken gauge theory, identical to the one proposed in sections 2.6.1 and 2.6.2.

2.7 The Minimal Standard Model

Combining the low energy information on the flavor symmetries with the observation of the top quark and the description of weak interactions by gauge bosons of a non-linearly realized $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y/\mathrm{U}(1)_Q$ and the unbroken $\mathrm{SU}(3)_C$ of QCD, results in the particles of the minimal SM lined up with their quantum number assignments in table 2.1. Instead of realizing $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y/\mathrm{U}(1)_Q$ non-linearly, it is also possible to introduce an elementary Higgs particle to break the electroweak symmetry.

Both options provide a consistent theoretical framework for calculations and are not ruled out by experiment. However, a linear realization with a light Higgs near 100 GeV appears to be favored by current experimental results (see figure 1.2).

As discussed at length in section 2.6, the low energy non-abelian charged currents require that among the equivalent effective theories describing the vector bosons, there is one in which the bosons are the non-abelian gauge bosons of a broken symmetry. Therefore the non-abelian couplings shown in figure 1.5 are unavoidable and deviations from the gauge theory predictions have to be consistent with the power counting rules of EFT, i. e. Naive Dimensional Analysis (NDA), as described in section A.3.5.

NDA is based on the hierarchy of two scales, here the Fermi scale $v \approx 246\,\mathrm{GeV}$ and $\Lambda_{\mathrm{EWSB}} = 4\pi v \approx 3.1\,\mathrm{TeV}$, with no intermediate scale. Therefore its predictions can be invalidated by the observation of a new physical threshold between v and Λ_{EWSB}. However, such a threshold would be observed and NDA is a valid tool for studying electroweak gauge bosons on the TeV scale, if no new particles will have been discovered until then.

Unless a new threshold is discovered or something very spectacular happens that would undermine most our present understanding of QFT, the deviations from the minimal SM must be on the order of

$$\frac{1}{16\pi^2} \approx 6.3 \cdot 10^{-3} . \tag{2.48}$$

From the predictions for the sensitivity at LEP2 on gauge boson couplings [18], we see that (2.48) will not be reached. Thus LEP2 will *not* be able to distinguish among models for EWSB, based on measurements of the gauge boson couplings alone. The direct search for a very light Higgs is more promising. This situation changes will change dramatically at the LC, due to the increased energy and luminosity (see section 5.1).

2.8 Precision Tests

The observation that low energy physics can be described by an effective field theory, does *not* mean that precision experiments can not access at least some aspects of the physics at higher scales through loop corrections [41]. If the low energy EFT has less symmetry than the EFT relevant at a higher energy scale, the parameters of the low energy EFT can be tested for the remnants of the high energy symmetry. The most celebrated example of this phenomenon at LEP1 are the radiative corrections to the ρ-parameter and the evidence for a custodial $\mathrm{SU}(2)_c$ symmetry [30] of the EWSB sector (see also section A.3.1).

At low energies, the relative strengths of the weak and neutral current couplings is a free parameter, in the current-current couplings (2.15). After the introduction of intermediate vector bosons, the two dimensionful parameters are interpreted as the ratio of dimensionless couplings and vector boson masses. Above the production threshold of the vector bosons, their masses can be measured independently from the strength of weak interactions. From the observed values, they appear to be related because the ρ-parameter

$$\rho = \frac{M_W^2}{\cos^2\theta_w M_Z^2} \tag{2.49}$$

is very close to unity, with the renormalization scheme dependent[8] deviations in the SM on the order of 1%. In (2.49), the weak mixing angle θ_w is defined

$$e = g\sin\theta_w \tag{2.50}$$

by the ratio of the left handed charged current couplings and the electromagnetic coupling.

Of course, $\rho = 1$ could be just a numerical accident, unless it can be explained by a symmetry. In the minimal SM, $\rho = 1$ is explained by the fact that all three components of the isospin-triplet charged gauge boson receive the *same* mass from the coupling to the EWSB sector, *before* the neutral component mixes with the hypercharge gauge boson. In the linear realization of the EWSB, the global $\mathrm{SU}(2)_c \otimes \mathrm{SU}(2)_L$ symmetry is equivalent to the SO(4) of the four components of the complex Higgs doublet. In a Higgs-less realization, the $\mathrm{SU}(2)_c \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y/\mathrm{SU}(2)_{C+L} \otimes \mathrm{U}(1)_Q$ has to be realized non-linearly [25]. Apparently, whatever the detailed dynamics of EWSB is, it respects the *custodial symmetry* relating W^+, W^-, and $W^3 = \cos\theta_w Z^0 + \sin\theta_w A$.

The most important deviations from $\rho = 1$ in the minimal SM depend on m_t (see section A.3.1). Therefore, the joint efforts of LEP1, LEP2, SLC and Tevatron in the measurement of M_Z, M_W, $\sin\theta_w$, and m_t have provided strict constraints on the symmetry of the EWSB sector. The TeV-scale physics probed in this way exceeds the direct physics reach of the contributing colliders (see figure 1.1) by an order of magnitude.

[8] There are mass dependent renormalization schemes that actually enforce $\rho = 1$ at one scale. The non-trivial observation is then that ρ does not "run away" from this value.

—3—
Matrix Elements

The quest for a theory of flavor demands precise calculations of high energy scattering processes in the framework of the SM and its extensions. At the Tevatron, the Large Hadron Collider (LHC), and at the LC, final states with many detected particles and with tagged flavor will be the primary handle for testing theories of flavor.

3.1 Three Roads from Actions to Matrix Elements

3.1.1 Manual Calculation

The traditional textbook approach to deriving amplitudes from actions are *manual calculations*, which can be aided by computer algebra tools. The time-honored method of calculating the squared amplitudes directly using trace techniques is no longer practical for today's multi particle final states and has generally been replaced by helicity amplitude methods (see e. g. [42, 43, 44]).

Manual calculations have the disadvantage of consuming a lot of valuable physicist's time, but can provide insights that are hidden by the other approaches discussed below. Regarding basic processes, where extremely fast and accurate predictions are required for numerical fits, manual calculations can provide efficient formulae that are still unrivaled.

3.1.2 Computer Aided Calculation

An increasingly popular technique is to use a well tested *library of basic helicity amplitudes* for the building blocks of Feynman diagrams and to construct the complete amplitude directly in the program in the form of function calls. A possible disadvantage is that the differential cross section is nowhere available as a formula, but the value of such a formula is limited anyway, since they can hardly be printed on a single page anymore.

3.1.3 Automatic Calculation

Fully *automatic calculations* are a further step in the same direction. The Feynman rules (or equivalent prescriptions) are no longer applied manually, but encoded algorithmically. This method will become more and more important in the future, but more work is needed for the automated construction of efficient unweighted event generators (see section 4.1.4) and for the automated factorization of collinear singularities in radiative corrections (see [45]).

3.2 Forests and Groves

Calculations of cross section with many-particle final states remain challenging despite all technical advances and it is of crucial importance to be able to concentrate on the important parts of the scattering amplitude for the phenomena under consideration.

In gauge theories, however, it is impossible to simply select a few signal diagrams and to ignore irreducible backgrounds. The same subtle cancellations among the diagrams in a gauge invariant subset that lead to the celebrated good high energy behavior of gauge theories such as the SM (see section 2.6), come back to haunt us if we accidentally select a subset of diagrams that is not gauge invariant. The results from such a calculation have *no* predictive power, because they depend on unphysical parameters introduced during the gauge fixing of the Lagrangian. It must be stressed that not all diagrams in a gauge invariant subset have the same pole structure and that a selection based on "signal" or "background" will not suffice.

The subsets of Feynman diagrams selected for any calculation must therefore form a *gauge invariant subset*, i. e. together they must already satisfy the Ward- and Slavnov-Taylor-identities to ensure the cancellation of contributions from unphysical degrees of freedom.

In abelian gauge theories, such as QED, the classification of gauge invariant subsets is straightforward and can be summarized by the requirement of inserting any additional photon into *all* connected charged propagators, as shown in figure 3.1. This situation is similar for gauge theories with simple gauge groups, the difference being that the gauge bosons are carrying charge themselves. For non-simple gauge groups like the SM, which even includes mixing, the classification of gauge invariant subsets is much more involved.

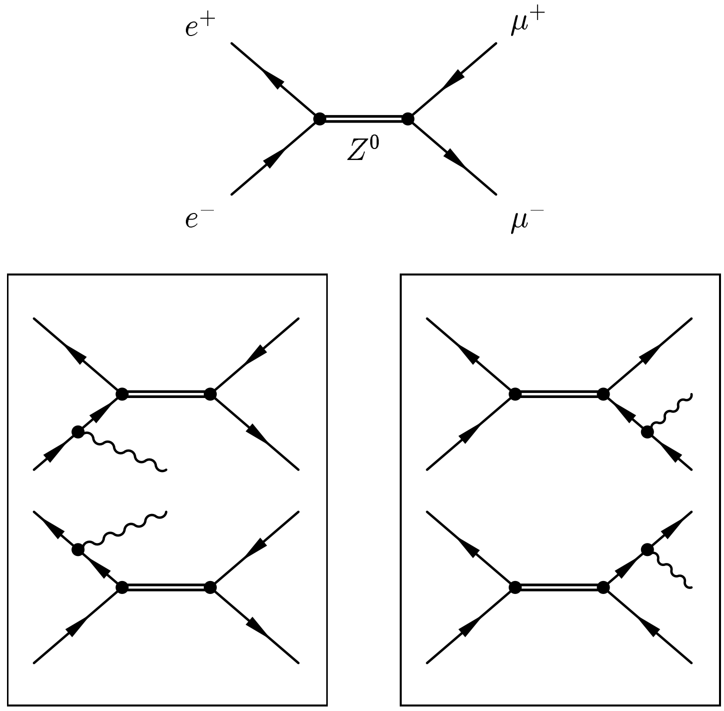


Figure 3.1: *The gauge invariant subsets in $e^+e^- \to \mu^+\mu^-\gamma$ are obtained by adding the photon in all possible ways to connected lines of charged particles.*

In early calculations, the classification of gauge invariant subsets in the SM has been performed in an ad-hoc fashion (see [46, 47]), which was later superseded by the explicit construction of *groves* [48], the smallest gauge invariant classes of tree Feynman diagrams in gauge theories. This construction is not restricted to gauge theories with simple gauge groups. Instead, it is applicable to gauge groups with any number of factors, which can even be mixed, as in the SM. Furthermore, it does not require a summation over complete multiplets and can therefore be used in flavor physics when members of weak isospin doublets (such as charm or bottom) are detected separately. The method constructs the smallest gauge invariant subsets and examples below illustrate that they are indeed smaller than those derived from looking at the final state alone [46, 47].

The method will probably also have applications in loop calculations. However, the published proofs use properties of tree diagrams and further research is required in this area.

In unbroken gauge theories, the permutation symmetry of external gauge quantum numbers can be used to subdivide the scattering amplitude corresponding to a grove further into gauge invariant sub-amplitudes (see [49] and references cited therein). In this decomposition, each Feynman diagram contributes to more than one sub-amplitude. It is not yet known how to perform a similar decomposition systematically in the SM, because the entanglement of gauge couplings and gauge boson masses complicates the structure of the amplitudes. The construction of the groves provides a necessary first step towards the solution of this important problem.

3.2.1 Gauge Cancellations

Even if the general arguments are well known, the explicit cancellations of unphysical contributions remain amazing in explicit perturbative calculations in gauge theories. The intricate nature of these cancellations complicates the development of systematic calculational procedures. Nevertheless, phenomenology needs numerically well behaved matrix elements and cross sections, which should be as compact as possible. The development of an automated procedure for obtaining compact expressions with explicit gauge cancellations remains a formidable challenge. The selection of gauge invariant subsets of Feynman diagrams described in the remainder of this section is one necessary step towards this goal.

As discussed in chapter 2, there is overwhelming phenomenological evidence that the strong and electroweak interactions are mediated by vector particles. A naive description of vector particles would assign to them four degrees of freedom or polarization states, of which only two (for massless particles) or three (for massive particles) can be physical. The only known theories that can consistently decouple the unphysical degrees of freedom for interacting fields are gauge theories (for massless particles) or equivalent to spontaneously broken gauge theories (for massive particles).

However, perturbative calculations require an explicit breaking of gauge invariance for technical reasons and the cancellation of unphysical contributions is not manifest in intermediate stages of calculations. The contribution of a particular Feynman diagram to a scattering amplitude depends in the gauge fixing procedure and has no physical meaning, in general.

Among the possible gauge fixing schemes, two complementary approaches can be identified: *unitarity gauges* and *renormalizable gauges.* A massive gauge boson in unitarity gauge has a propagator

$$-\frac{i}{p^2-M^2+i\epsilon}\left(g_{\mu\nu}-\frac{p_\mu p_\nu}{M^2}\right) \tag{3.1}$$

and unitarity of the time evolution can be proved straightforwardly, because only physical degrees of freedom propagate and the condition $\partial_\mu V^\mu = 0$ is maintained. Unfortunately, the longitudinal part of the propagator (3.1) does not fall off with high energy and momentum. This spoils the power counting for renormalization and, even worse, violates tree-level unitarity at high energies, because the scattering amplitudes for longitudinal vector bosons do not fall off rapidly enough to satisfy the optical theorem at high energies.

On the other hand, the family of R_ξ-gauge fixings results in propagators that propagate unphysical degrees of freedom

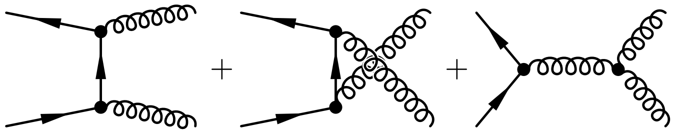
$$-\frac{i}{p^2 - M^2 + i\epsilon}\left(g_{\mu\nu} - (1-\xi)\frac{p_\mu p_\nu}{p^2 - \xi M^2 + i\epsilon}\right) \tag{3.2}$$

but with good high energy behavior and standard power counting even for the longitudinal components. The cancellation of the unphysical contributions in scattering matrix elements turns out to be a formidable combinatorial problem [50].

The second approach is superior for formal proofs to all orders (see [51] and references therein), because it allows to treat the two complicated problems independently. In a first step, the theory is defined using a Bolgoliubov-Parasiuk-Hepp-Zimmermann (BPHZ) subtraction procedure [52], which violates gauge invariance and unitarity, but allows to prove the Quantum Action Principle (QAP) [53]. The QAP states that the classical symmetries of the Lagrangian density L are respected by the effective action Γ, upto a *local* operator of dimension five (see also [54]). Subsequently, one can prove by an inspection of all operators of dimension five, that the BRS invariance of the effective action can be restored by *finite* and *local* counterterms of dimension four, provided that the coefficient of the Adler-Bell-Bardeen-Jackiw (ABBJ) anomaly vanishes. Finally, the unitarity of the BRS invariant theory is proven with the quartet-mechanism [33].

This two-step procedure is transparent because it decouples the combinatorics of renormalized perturbative power counting from the combinatorics of gauge invariance and BRS invariance. Once the QAP is established, BRS invariance is reduced to a problem in finite-dimensional cohomology. On the other hand, controlling the high energy behavior of the unitarity gauge propagator (3.1) and the combinatorics of Ward identities simultaneously is prohibitively complicated, even in low orders of the loop expansion.

However, this decoupling of kinematical structure and gauge structure, which makes the abstract argument transparent, is not useful in concrete calculations, because the abstract arguments apply only to the sum of all diagrams and shed little light on how the gauge cancellations “work”. One particular problem is that the Ward identities relate diagrams with different pole structure. For example, in $q\bar{q} \to gg$

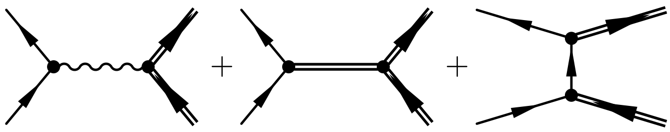


(3.3)

the numerator factors cancel parts of denominators and the Ward identity is satisfied only by the sum and not by individual diagrams.

Therefore, the physical amplitudes are determined by an intricate web of kinematical structure and gauge structure, that is very hard to disentangle. Nevertheless, the identification of partial sums of Feynman diagrams that are gauge invariant by themselves is of great practical importance for at least two reasons. Firstly, it is more economical to spent the time improving Monte Carlo statistics for the important pieces of the amplitude than to calculate the complete amplitude all the time. This requires the identification of the smallest gauge invariant part of the amplitude that contains the important pieces. Secondly, different parts of the amplitude are characterized by different scales (e. g. s or t) for radiative corrections. The different scales can only be used if the parts correspond to separately gauge invariant pieces.

The practical consequence of these arguments for LEP2 physics is that each individual diagram in the $e^+e^- \to W^+W^-$ amplitude



(3.4)

has a bad high energy behavior [40] and only the inclusion of subtle interferences in the sum results in good high energy behavior[1].

The on-shell pair production (3.4) is unfortunately not the observed process. This causes two problems: the finite width necessitates Dyson-resummation and higher orders violate gauge invariance [56]. Furthermore, the observed final state consists of many observed particles with tagged flavor. Therefore there are many tree-level background diagrams that contribute to the same process. For this reason, the calculations are complicated and it is important to identify the important parts.

3.2.2 Flavor Selection Rules

The interesting flavor physics at the LHC and at the LC is different from QCD with light quarks because the gauge symmetry is broken spontaneously and

[1] The good high energy behavior is realized separately for the $SU(2)_L$ and $U(1)_Y$ parts of the amplitude [55], but this requires breaking up diagrams.

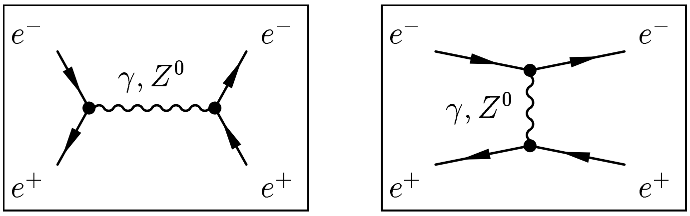


Figure 3.2: *Gauge invariant subsets of Feynman diagrams in Bhabha scattering.*

the gauge group is not simple. Furthermore, the factors are mixed according to

$$\begin{pmatrix} Z^0 \\ A \end{pmatrix} = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} W^3 \\ B \end{pmatrix} \tag{3.5}$$

and the combinatorics are consequently more intricate. Finally, the flavor of leptons and heavy quarks is tagged and the flavor transformations do not commute with the gauge transformations and no summation over complete gauge multiplets is possible. For these reasons, the QCD recursion relations or string inspired methods [49] are not directly applicable.

One method of identifying gauge invariant subsets is to identify subsets that contribute to a particular final state [46, 47]. For example, there are ten tree diagrams contributing to $e^+e^- \to \mu^-\bar\nu_\mu u\bar d$ and therefore the corresponding subset of the 20 diagrams in $e^+e^- \to e^-\bar\nu_e u\bar d$ and its complement must be both gauge invariant by themselves. The subsets that can be derived in this way are in general not minimal, but this example shows that selection rules of flavor symmetries that commute with the gauge group appear to be a useful tool.

The simplest example of these flavor selection rules is provided by Bhabha scattering: the s-channel and the t-channel diagram in figure 3.2 have to be gauge-invariant separately, because both $e^+e^- \to \mu^+\mu^-$ and $e^+\mu^- \to e^+\mu^-$ are physical processes that must give gauge invariant amplitudes by themselves. The first process has only the s-channel diagram and the second only the t-channel diagram. As a result, we see that conserved currents of (real or fictitious) horizontal (i. e. generation) symmetries can serve as a tool to separate gauge invariant classes of Feynman diagrams.

A less trivial example is provided by the three separately gauge invariant sets in $ud \to ud$ scattering shown in figure 3.3. The separate gauge invariance of the gluon exchange diagram is obvious because the QCD generators commute with the electroweak generators and the strong coupling can be

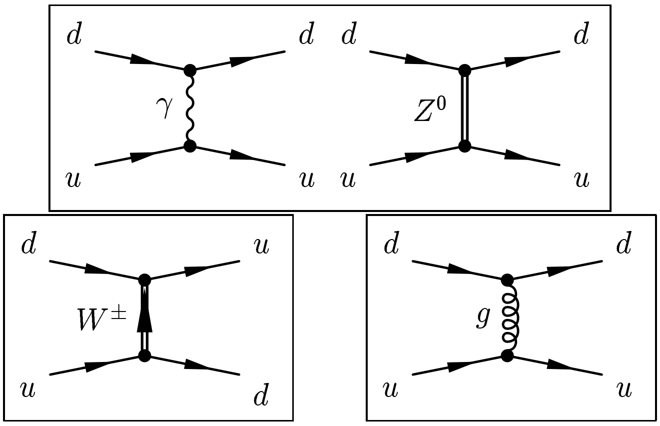


Figure 3.3: *Gauge invariant subsets of Feynman diagrams in $ud \to ud$ scattering.*

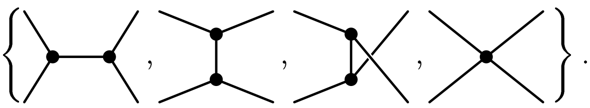
switched on and off without violating gauge invariance. The charged current diagram is separately gauge-invariant, because we may assume that the CKM mixing matrix is diagonal and then the charged current diagonal is absent in $us \to us$ scattering, which is related to $ud \to ud$ by a horizontal symmetry that commutes with the gauge group.

These example in $2 \to 2$ scattering are not useful by themselves, unless we can handle the combinatorics in the more complicated realistic applications.

3.2.3 Forests

In order to develop tools for understanding the combinatorics, we have to take a step back from the intricacies of gauge theories and study unflavored scalar ϕ^3- and ϕ^4-theory. We assume that no ϕ^5 or higher vertices are required by phenomenology. In principle, the combinatorial methods describe below can be extended to non-renormalizable theories. However, the required arguments would have be more involved, without providing additional insight into the gauge sector. Since there are no selection rules, the diagrams S_1, S_2, and S_3 in

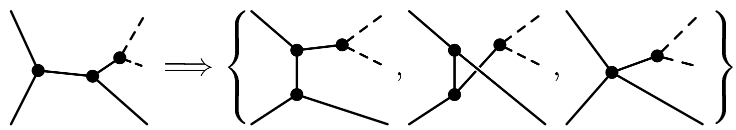
$$T_4 = \{t_4^{S,1}, t_4^{S,2}, t_4^{S,3}, t_4^{S,4}\} = \tag{3.6}$$



must have the same coupling strength to ensure crossing invariance. If there are additional symmetries, as in the case of gauge theories, the coupling of $t_4^{S,4}$ will also be fixed relative to $t_4^{S,1}$, $t_4^{S,2}$, and $t_4^{S,3}$.

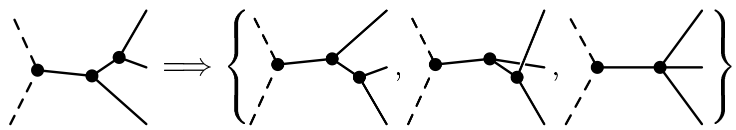
Henceforth, we will call each exchange $t \leftrightarrow t'$ of two members of the set T_4 of all tree graphs with four external particles an *elementary flip*. The elementary flips define a trivial relation $t \circ t'$ on T_4, which is true, if and only if t and t' are related by a flip. The relation $\circ$ is trivial on T_4 because *all* pairs are related by an elementary flip.

However, the elementary flips in T_4 induce *flips* in T_5 (the set of all tree diagrams with five external particles) if they are applied to an arbitrary four particle subdiagram



(3.7a)

Obviously, there is more than one element of T_4 embedded in a particular $t \in T_5$ and the same diagram is member of other quartets, e. g.



(3.7b)

There are 15 five-point tree diagrams in ϕ^3 and inspection shows that only four other inequivalent diagrams can be reached from any diagram by flips of a four-point subdiagram. Thus there must be a non-trivial mathematical structure on the set T_5 with the relation $\circ$. This structure can be visualized by the graph in figure 3.4. Similarly, there are 25 five-point tree diagrams in combined ϕ^3 and ϕ^4 and at most six can be reached from any diagram by a flip.

In the same way, the trivial relation on T_4 has a non-trivial natural extension to the set T_n of all n-point tree diagrams: $t \circ t'$ is true if and only if t and t' are identical up to a single flip of a four-point subdiagram

$$t \circ t' \Longleftrightarrow \exists t_4 \in T_4, t'_4 \in T_4 : t_4 \circ t'_4 \wedge t \setminus t_4 = t' \setminus t'_4 \tag{3.8}$$

Note that this relation is not transitive and therefore not an equivalence relation. Instead, this relation allows us to view the elements of T_n as the vertices of a graph F_n, where the edges of the graph are formed by the pairs of diagrams related by a single flip

$$F_n = \left\{ (t,t') \in T_n \times T_n \middle| t \circ t' \right\}. \tag{3.9}$$

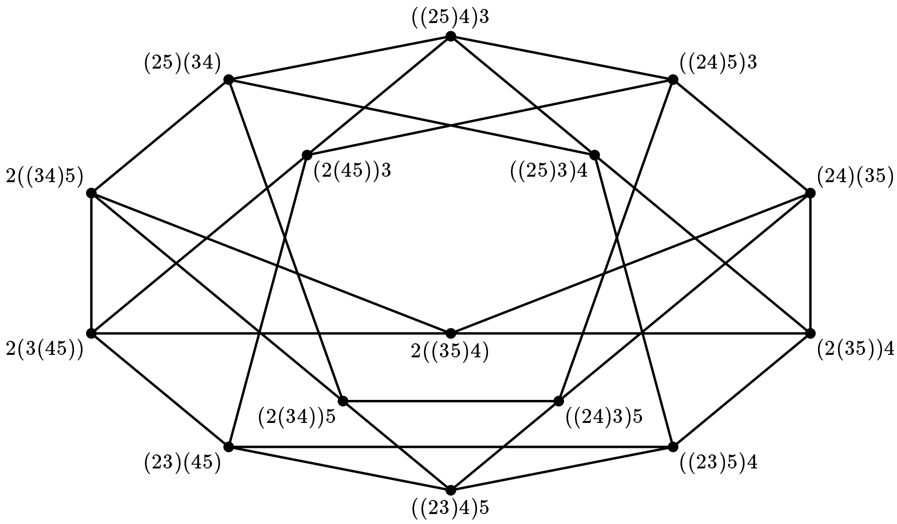


Figure 3.4: *The forest F_5 of the 15 five-point tree diagrams in unflavored ϕ^3-theory. The diagrams are specified by fixing vertex 1 and using parentheses to denote the order in which lines are joined at vertices (see the footnote on page 41 for examples).*

To avoid confusion, we will refer to graph F_n as *forest* and to its vertices as Feynman *diagrams*.

The most important property of the forests for our applications is

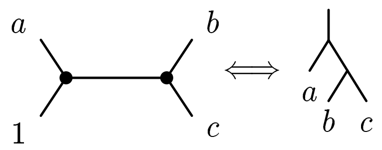
Theorem 1 *The unflavored forest F_n is connected for all n.*

which is easily proved by mathematical induction on n. Indeed, all members of T_{n+1} can be derived from T_n by connecting a new external line to a propagator (forming a new ϕ^3-vertex) or an existing ϕ^3-vertex (transforming it into ϕ^4-vertex). Obviously, two such insertions are related by a finite number of flips, which proves the induction step. This theorem shows that it is possible to construct all Feynman diagrams by visiting the nodes of F_n along successive applications of the flips in (3.6).

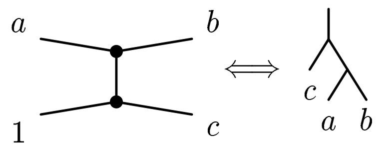
Already the simplest non-trivial example of a unflavored ("vanilla") forest, the 15 tree diagrams with five external particles in unflavored ϕ^3-theory, as shown in figure 3.4, displays an intriguing symmetry structure: there are 120 permutations of the vertices of F_5 that leave this forest invariant. The automorphism group $\mathbf{Aut}(F_5)$ is therefore with 120 elements much larger than one would expect. The classification of vanilla forests is probably an interesting mathematical problem, but without obvious physical applications. However, as we will see below, not all flips are created equal and only some flips are required by gauge invariance, leading to a partition of the forest into gauge invariant subsets, called *groves* [48].

Vanilla Trees

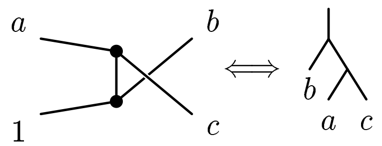
The enumeration of tree diagrams is simplified considerably if we can find a *unique* representative for each topological equivalence class [57]. One such representation is defined by picking one external line as the root and drawing the reminder of the diagram as a mathematical tree. The edges at each node are sorted according to the index of the external lines beneath each node. The sorting lifts the remaining degeneracy and there remains only one representative for each topological equivalence class. To be specific, here is how the members of T_4 are represented



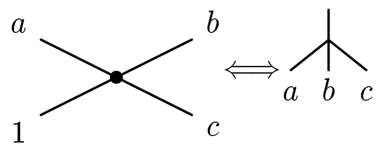
(3.10a)



(3.10b)

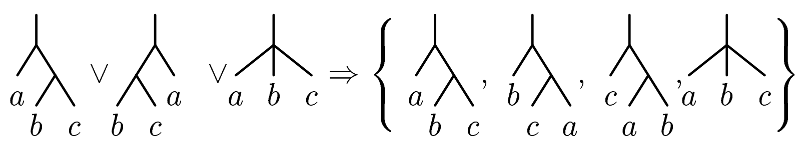


(3.10c)



(3.10d)

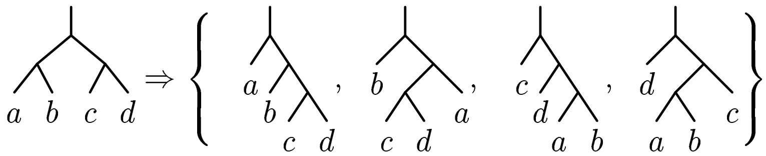
In this representation[2], the flips are equivalent to a recursive *tree pattern matching*



(3.11a)

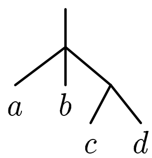
[2]Instead of the graphical representation used in (3.10—3.12), a notation based on parenthesis has been used in figure 3.4. In this less intuitive but more concise notation, the four trees in (3.10) are denoted by $a(bc)$, $c(ab)$, $b(ac)$, and abc, respectively. The matching in (3.11b) is written $(ab)(cd) \Rightarrow \{a(b(cd)), b((cd)a), c(d(ab)), d((ab)c)\}$.

where each of a, b, c can be either a complete subtree or a single leaf. In these pattern matchings, a subtle special case in which two eligible vertices meet, has to be taken care of



(3.11b)

in order not to miss a branch. However, cases like



(3.12)

need *no* special rules because the root vertex will only be expanded and never be contracted with another vertex.

This unique representation of tree diagrams provides an efficient equality predicate and the recursive pattern matching provides a function calculating the neighbors

$$\phi(t) = \{t' \in T | t' \circ t\} \tag{3.13}$$

of any tree t in the forest. Using these two ingredients, the construction of the forest is a well defined mathematical problem with efficient textbook solutions, known as *depth first search* and *breadth first search*. Essentially, these solutions consist of recursively following the edges to the neighbors until all nodes have been visited, keeping track of the visited nodes to avoid loops.

If the forest is connected (from theorem 1, the vanilla forests are known to be connected) and a single tree is known, the construction of the forest is also an efficient algorithm for generating all Feynman tree diagrams.

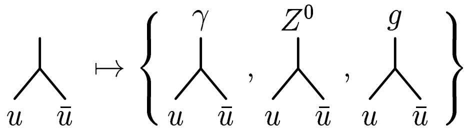
3.2.4 Flavored Forests

In physics applications, we have to deal with multiple flavors of particles. Therefore we introduce *flavored forests*, where the admissibility of elementary flips $t \circ t'$ depends on the four particles involved through the Feynman rules for the vertices in t_4 and t'.

In order to simplify the combinatorics when counting diagrams for theories with more than one flavor, we will below treat *all* external particles as outgoing. The physical amplitudes are obtained later by selecting the incoming particles in the end. Ward identities, etc. will be proved for the latter physical amplitudes, of course.

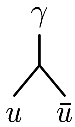
Flavored trees are derived from vanilla trees by adding a label to each edge. Feynman rules are then used at each node to identify valid flavor combinations.

All flavored trees can be generated from vanilla trees (which correspond to "topologies") by a recursive tensor product of elementary alternatives. E. g.



(3.14)

The representation described in section 3.2.3 allows to use efficient tree homomorphisms for the necessary manipulations. In particular, the calculation of the algebraic expressions corresponding to flavored trees is a tree homomorphism:



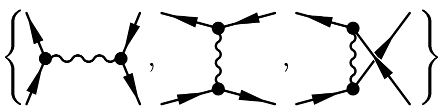
$$\mapsto \bar{u}(p_1)\gamma_\mu u(p_2)\,. \tag{3.15}$$

If there were no Feynman rules, the flavored forests would be connected just like the vanilla forests. However, the Feynman rules typically prohibit most of the possible flavor assignments and it turns out that the remaining vertices form more than one connected component.

3.2.5 Groves

The construction of the groves is based on the observation that the flips in gauge theories fall into two different classes: the *flavor flips* among

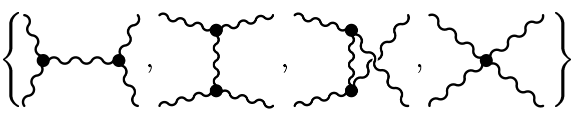
$$\{t_4^{F,1}, t_4^{F,2}, t_4^{F,3}\} =$$



(3.16)

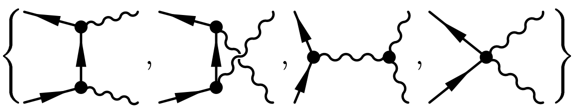
which involve four matter fields that carry gauge charge and possibly additional conserved quantum numbers and the *gauge flips* among

$$\{t_4^{G,1}, t_4^{G,2}, t_4^{G,3}, t_4^{G,4}\} =$$



(3.17a)

and

$\{t_4^{G,5}, t_4^{G,6}, t_4^{G,7}, t_4^{G,8}\} =$  (3.17b)

which also involve external gauge bosons (the diagram $t_4^{G,8}$, is only present for scalar matter fields that appear in extensions of the SM: SUSY partners, leptoquarks, etc.). In gauge theories with more than one factor, like the SM, the gauge flips are extended in the obvious way to include all four-point functions with at least one gauge-boson. Commuting gauge group factors lead to separate sets, of course. In spontaneously broken gauge theories, the Higgs and Goldstone boson fields contribute additional flips, in which they are treated like gauge bosons. Ghosts can be ignored at tree level.

As we have seen for the examples of Bhabha-scattering and $ud \to ud$ scattering in section 3.2.2, the flavor flips (3.16) are special because they can be switched off without spoiling gauge invariance by introducing a horizontal symmetry that commutes with the gauge group. Such a horizontal symmetry is similar to the replicated generations in the SM, but if three generations do not suffice, it can also be introduced artificially. This observation suggests to introduce two relations:

$$t \bullet t' \iff t \text{ and } t' \text{ related by a \textit{gauge} flip} \tag{3.18a}$$

and

$$t \circ t' \iff t \text{ and } t' \text{ related by a \textit{flavor or gauge} flip.} \tag{3.18b}$$

These two relations define two different graphs with the same set $T(E)$ of all Feynman diagrams as vertices:

$$F(E) = \left\{(t,t') \in T(E) \times T(E) \middle| t \circ t'\right\} \tag{3.19a}$$

and

$$G(E) = \left\{(t,t') \in T(E) \times T(E) \middle| t \bullet t'\right\}. \tag{3.19b}$$

As we have seen above in Bhabha-scattering, it is in general not possible to connect all pairs of diagrams in $G(E)$ by a series of gauge flips. Thus there will be more than one connected component. For brevity, we will continue to denote the *flavor forest* $F(E)$ as the *forest* of the external state E and we will denote the connected components $G_i(E)$ of the *gauge forest* $G(E)$ as the *groves* of E. Since $t \bullet t' \Rightarrow t \circ t'$, we have $\bigcup_i G_i(E) = G(E) \subseteq F(E)$, i. e. the groves are a *partition* of the gauge forest.

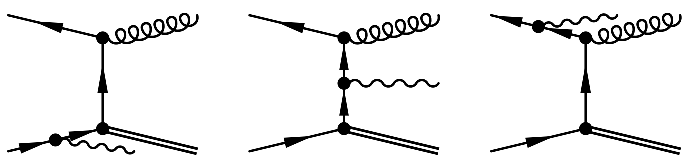


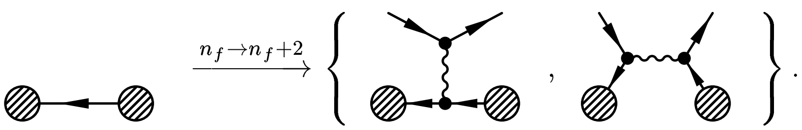
Figure 3.5: *The three diagrams in* $q\bar{q} \to gZ^0\gamma$ *are connected by successive gauge flips of subdiagrams like (3.17b). If the decays of the* Z^0 *were included, bremsstrahlung from the decay products would form a separate grove, because no gauge flip will pass the photon across the uncharged* Z^0*.*

Theorem 2 *The forest* $F(E)$ *for an external state* E *consisting of gauge and matter fields is connected if the fields in* E *carry no conserved quantum numbers other than the gauge charges. The groves* $G_i(E)$ *are the minimal gauge invariant classes of Feynman diagrams.*

By construction, the theorem is true for the four-point diagrams and we can use induction on the number of external matter fields and gauge bosons. Since the matter fields are carrying conserved charges, they can only be added in pairs.

If we add an additional external gauge boson to a gauge invariant amplitude, the diagrammatical proof of the Ward and Slavnov-Taylor identities in gauge theories requires us to sum over all ways to attach a gauge boson to connected gauge charge carrying components of the Feynman diagrams. However, the gauge flips are connecting pairs of neighboring insertions and can be iterated along gauge charge carrying propagators. Therefore no partition of the forest $F(E)$ that is finer than the groves $G_i(E)$ can preserve gauge invariance. Two examples are shown in figures 3.5 and 3.6.

If we add an additional pair of matter fields to a gauge invariant amplitude, we have to consider two separate cases



(3.20)

If the new flavor does not already appear among the other matter fields, the only way to attach the pair is through a gauge boson, as in the left diagram in the braces in (3.20). If the new flavor is already present, we can also break up a matter field propagator and connect the two parts of the diagram with a new gauge propagator. Since it is always possible to introduce a new flavor, either physical or fictitious, without breaking gauge invariance, these

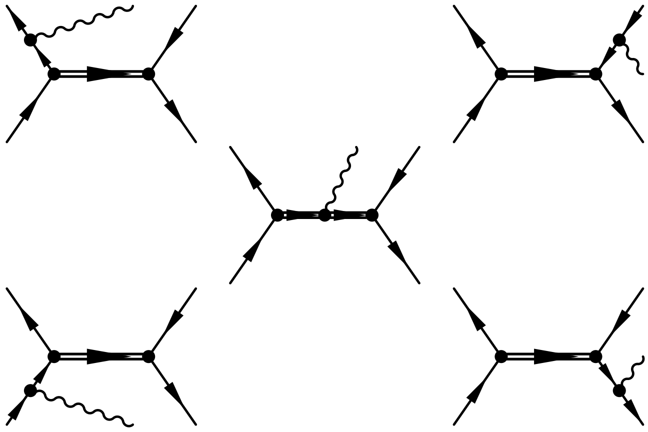


Figure 3.6: *All five diagrams in* $u\bar{d} \to c\bar{s}\gamma$ *are in the same grove, because they are connected by gauge flips passing through the diagram in the center. In contrast, in* $e^+e^- \to \mu^+\mu^-\gamma$ *(see figure 3.1), the center diagram is missing and there are two separate groves.*

cases fork off separately gauge invariant classes every time we add a new pair of matter fields. On the other hand, the cases in (3.20) are related by a flavor flip. Therefore $F(E)$ remains connected, the $G_i(E)$ are separately gauge invariant and the proof is complete.

Earlier attempts [46, 47] have used physical final states as a criterion for identifying gauge invariant subsets. We have already shown that the groves are minimal and therefore never form a more coarse partition than the one derived from a consideration of the final states alone. Below we shall see examples where the groves do indeed provide a strictly finer partition.

In a practical application, one calculates the groves for the interesting combinations of gauge quantum numbers, such as weak isospin and hypercharge in the SM, using an external state where all other quantum numbers are equal, since the flavor forest would otherwise be disconnected. The combinatorics is simplified by treating all particles as outgoing. The physical amplitude is then obtained by selecting the groves that are compatible with the other quantum numbers of the process under consideration. Concrete examples are considered in the next section.

If only a few flavor combinations are interesting, it can be more convenient in practical applications to generate all diagrams for the external states conventionally and to partition the disconnected flavor forest. The resulting groves will be the same, of course.

3.2.6 Application

In table 3.1, we list the groves for all processes with six external massless fermions in the SM, with and without single photon bremsstrahlung, without QCD and CKM mixing. We could also include fermion masses, QCD, and CKM mixing within the same formalism, but the table would have to be much larger, because additional gluon, Higgs and Goldstone diagrams appear, depending on whether the fermions are massive or massless, colored or uncolored. In the table, cases with identical $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ quantum numbers are listed only once and cases with different T_3 and Y are listed separately only if the vanishing of the electric charge removes diagrams from a grove. E. g. $e^+e^-e^+e^-e^+e^-$ is identical to $u\bar{u}u\bar{u}u\bar{u}$ and not listed separately.

The familiar non-minimal gauge invariant classes for $e^+e^- \to f_1\bar{f}_2f_3\bar{f}_4$ [46] are included in table 3.1 as special cases. The LEP2 WW-classes CC09, CC10, and CC11 are immediately obvious. As a not quite so obvious example, the process $e^+e^- \to \mu^+\mu^-\tau^+\tau^-$ has the same $\mathrm{SU}(2)_L$ quantum numbers

Table 3.1: *The groves for all processes with six external massless fermions and bremsstrahlung in the SM (without QCD). All particles are considered as outgoing.*

external fields (E)	#diagrams	groves
$u\bar{u}u\bar{u}u\bar{u}$	144	$18\cdot\mathbf{8}$
$u\bar{u}u\bar{u}u\bar{u}\gamma$	1008	$18\cdot\mathbf{24}+36\cdot\mathbf{16}$
$u\bar{u}u\bar{u}d\bar{d}$	92	$4\cdot\mathbf{11}+6\cdot\mathbf{8}$
$u\bar{u}u\bar{u}d\bar{d}\gamma$	716	$4\cdot\mathbf{95}+6\cdot\mathbf{24}+12\cdot\mathbf{16}$
$\ell^+\ell^-u\bar{u}d\bar{d}$	35	$1\cdot\mathbf{11}+3\cdot\mathbf{8}$
$\ell^+\ell^-u\bar{u}d\bar{d}\gamma$	262	$1\cdot\mathbf{94}+3\cdot\mathbf{24}+6\cdot\mathbf{16}$
$\ell^+\nu d\bar{u}d\bar{d}$	20	$2\cdot\mathbf{10}$
$\ell^+\nu d\bar{u}d\bar{d}\gamma$	152	$2\cdot\mathbf{76}$
$\ell^+\ell^-\ell^+\nu d\bar{u}$	20	$2\cdot\mathbf{10}$
$\ell^+\ell^-\ell^+\nu d\bar{u}\gamma$	150	$2\cdot\mathbf{75}$
$\ell^+\nu\ell^-\bar{\nu}d\bar{d}$	19	$1\cdot\mathbf{9}+2\cdot\mathbf{4}+1\cdot\mathbf{2}$
$\ell^+\nu\ell^-\bar{\nu}d\bar{d}\gamma$	107	$1\cdot\mathbf{59}+2\cdot\mathbf{12}+2\cdot\mathbf{8}+2\cdot\mathbf{4}$
$\ell^+\nu\ell^-\bar{\nu}\ell^+\ell^-$	56	$4\cdot\mathbf{9}+4\cdot\mathbf{4}+2\cdot\mathbf{2}$
$\ell^+\nu\ell^-\bar{\nu}\ell^+\ell^-\gamma$	328	$4\cdot\mathbf{58}+4\cdot\mathbf{12}+4\cdot\mathbf{8}+4\cdot\mathbf{4}$
$\ell^+\nu\ell^-\bar{\nu}\nu\bar{\nu}$	36	$4\cdot\mathbf{6}+6\cdot\mathbf{2}$
$\ell^+\nu\ell^-\bar{\nu}\nu\bar{\nu}\gamma$	132	$4\cdot\mathbf{26}+2\cdot\mathbf{6}+4\cdot\mathbf{4}$
$\nu\bar{\nu}\nu\bar{\nu}\nu\bar{\nu}$	36	$18\cdot\mathbf{2}$

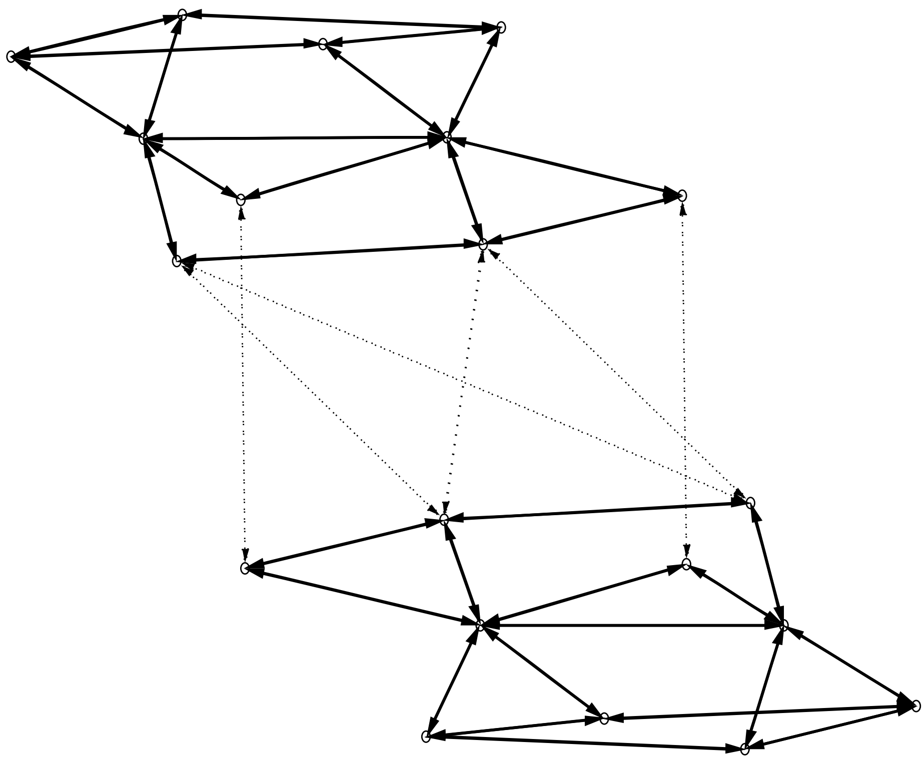


Figure 3.7: *The two groves in the CC20 process* $e^+e^- \to e^-\bar{\nu}_e u\bar{d}$. *Gauge flips are drawn a solid lines and flavor flips are drawn as dotted lines. These groves can be found in table 3.1 as* $\ell^+\ell^-\ell^+\nu d\bar{u}$.

as $u\bar{u} \to u\bar{u}u\bar{u}$. We can read off table 3.1 that, in the case of identical pairs, there are 18 groves, of 8 diagrams each. If all three pairs are different, the number of groves has to be divided by 3!, because we are no longer free to connect the three particle-antiparticle pairs arbitrarily. Thus there are 24 diagrams contributing to the process $e^+e^- \to \mu^+\mu^-\tau^+\tau^-$ and they are organized in three groves of 8 diagrams each. Any diagram in a grove can be reached from the other 7 by exchanging the vertices of the gauge bosons on one fermion line and be exchanging Z^0 and γ. Since there are no non-abelian couplings in this process, the separate gauge invariance of each grove could also be proven as in QED, by varying the hypercharge of each particle: $A \propto A_1 \cdot q_e^2 q_\mu q_\tau + A_2 \cdot q_e q_\mu^2 q_\tau + A_3 \cdot q_e q_\mu q_\tau^2$.

As is well known, the diagrams for the CC20 process $e^+e^- \to e^-\bar{\nu}_e u\bar{d}$ fall into two separately gauge invariant classes, depicted in figure 3.7 (including processes related by crossing symmetry). Figure 3.8 shows two views of the

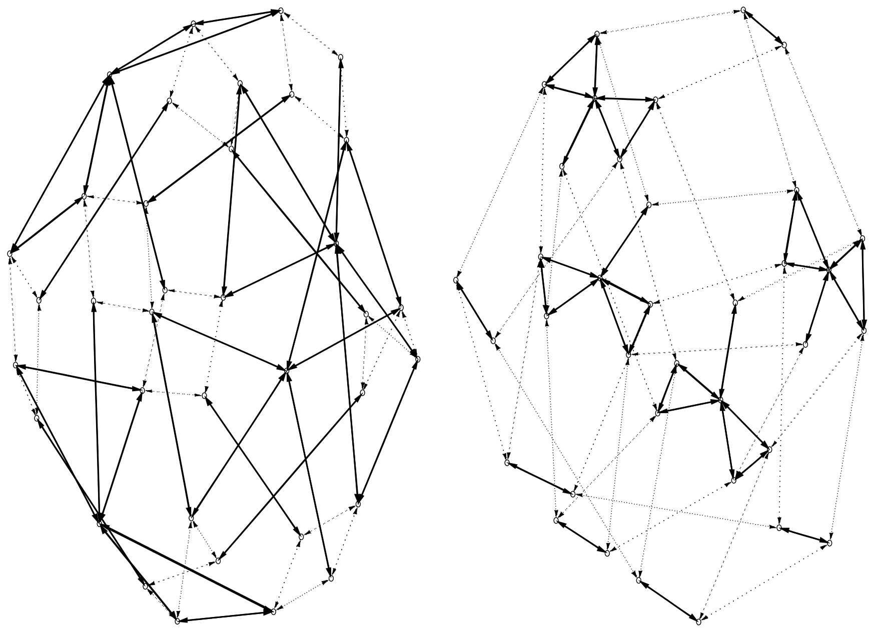


Figure 3.8: *Two views of $e^+e^- \to \nu_e\bar{\nu}_e\nu_e\bar{\nu}_e$. Gauge flips are drawn a solid lines and flavor flips are drawn as dotted lines.*

groves for $e^+e^- \to \nu_e\bar{\nu}_e\nu_e\bar{\nu}_e$ (including processes related by crossing symmetry). The left hand side emphasizes the four-fold permutation symmetry of flavor flips, while the right hand side emphasizes the groups of six diagrams forming a single grove. The process $e^+e^- \to \nu\bar{\nu}\nu\bar{\nu}\gamma$ is not yet the most pressing physics problem, but small enough to produce the nice picture shown in figure 3.9. The decomposition $4\cdot\mathbf{26}+2\cdot\mathbf{6}+4\cdot\mathbf{4}$ is clearly visible. Finally, figure 3.10 displays the forest for the process $\gamma\gamma \to u\bar{d}d\bar{u}$ in the SM. The grove in the center consists of the 31 diagrams with charged current interactions (the set CC21 of [47]). The four small groves of neutral current interactions are only connected with the rest of the forest through the charged current grove. The automorphism group of this forest has 128 elements.

The groves can now be used to select the Feynman diagrams to be calculated by other means. However, we note that it is also possible to calculate the amplitude with little additional effort already during the construction of the groves by keeping track of momenta and couplings in the diagram flips.

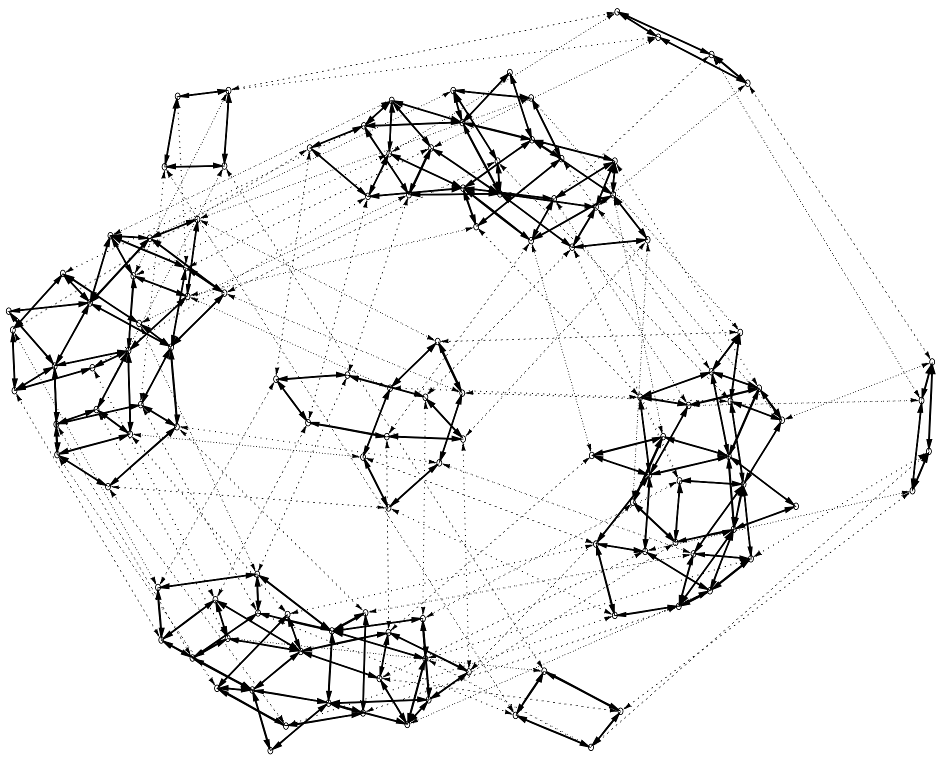


Figure 3.9: *The decomposition* $4 \cdot \mathbf{26} + 2 \cdot \mathbf{6} + 4 \cdot \mathbf{4}$ *of* $e^+e^- \to \nu\bar{\nu}\nu\bar{\nu}\gamma$*. Gauge flips are drawn a solid lines and flavor flips are drawn as dotted lines.*

3.2.7 Automorphisms

The forests and groves that we have studied appear to be very symmetrical in the neighborhood of any vertex. However, the global connection of these neighborhoods is twisted, which makes it all but impossible to draw the graphs in a way that makes these apparent symmetries manifest (see figure 3.8).

Nevertheless, one can turn to mathematics (in particular to computational and constructive group theory [58]) and construct the automorphism groups $\mathbf{Aut}(F(E))$ and $\mathbf{Aut}(G_i(E))$ of the forest $F(E)$ and the groves $G_i(E)$, i. e. the group of permutations of vertices that leave the edges invariant. These groups turn out to be larger than one might expect. For example, the group of permutations of the 71 vertices of the forest $F(\gamma\gamma \to u\bar{d}d\bar{u})$ in figure 3.10, that leave the edges invariant, has 128 elements. Similarly, the automorphism group of the forest in figure 3.4 has $120 = 5!$ elements. The study of these groups and their relations might enable us to construct gauge

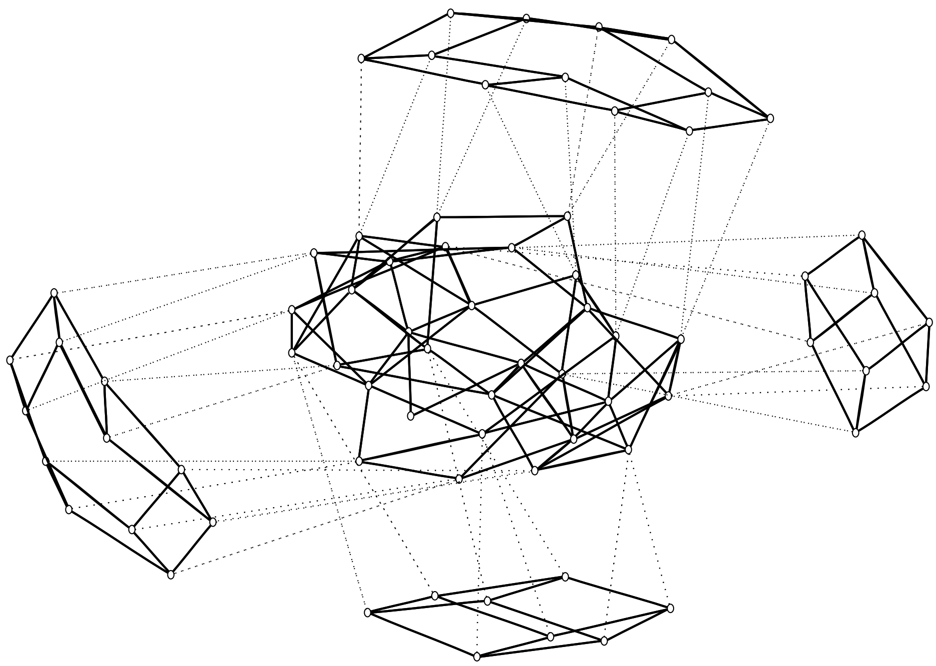


Figure 3.10: *The forest of size 71 for the process* $\gamma\gamma \to u\bar{d}d\bar{u}$ *in the SM (without QCD, CKM mixing and masses in unitarity gauge) with one grove of size 31, two of size 12 and two of size 8. Solid lines represent gauge flips and dotted lines represent flavor flips.*

invariant subsets directly.

3.2.8 Advanced Combinatorics

Even pure gauge forests can provide new insights into the combinatorics of gauge cancellations among Feynman diagrams. Consider the process $gg \to ggg$: there are 25 diagrams (all in the same grove, of course). 15 of them have three triple gluon vertices and the remaining 10 diagrams have one triple and one quadruple vertex each. The sub-forest of the 15 diagrams with triple vertices is equivalent to the one shown in figure 3.4. The 30 endpoints of the gauge flips starting from the 10 diagrams with a quadruple vertex cover this sub-forest with 15 nodes exactly twice.

Therefore, the gauge cancellations do *not* follow from a simple recursion of the gauge cancellations in $gg \to gg$, where the sub-forest of diagrams with triple couplings is covered exactly once. Unfortunately, the symmetries in $gg \to gggg$ are much more complicated. Nevertheless, the automorphism groups of groves should be studied in order to understand the gauge cancellations better.

3.2.9 Loops

The number of loops is invariant under flips, as can be seen from Euler's formula. However, some of the proofs above have made explicit use of properties of tree diagrams. Therefore further research is required in this area.

—4— Phase Space

Not from the stars do I my judgment pluck,
And yet methinks I have astronomy
[William Shakespeare: *Sonnet XIV*]

The calculation of matrix elements using the methods outlined in chapter 3 is but the first step towards a comparison of theoretical predictions with experiment. In the second step, the differential cross section obtained from squaring the amplitude has to be integrated in the region of phase space corresponding to the detector. While this step is common to all observables, it remains highly non-trivial and will be discussed in detail in this chapter.

In our application, the integration of a function f with a given measure μ on a manifold M

$$I(f) = \int_M \mathrm{d}\mu(p)\, f(p) \tag{4.1}$$

is complicated by three facts

1. the manifold of multi particle phase space and the corresponding Lorentz invariant measure

   $$\mathrm{d}\mu(p) = \delta^4(\textstyle\sum_n k_n - P) \prod_n \mathrm{d}^4 k_n\, \delta(k_n^2 - m_n^2) \tag{4.2}$$

   is non-trivial for three particles in the final state and analytically intractable for more than three massive particles,

2. the function is given by a squared matrix element multiplied by kinematical cuts or acceptances

   $$f(p) = |T(k_1, k_2, \ldots)|^2 \cdot C(k_1, k_2, \ldots) \tag{4.3}$$

   and usually ill-behaved. Integrable singularities are common in theories with massless particles and fluctuations over many orders of magnitude are the norm rather than the exception,

3. in realistic applications, the acceptances and efficiencies in $C(k_1, k_2, \dots)$ are almost never given as analytical functions. Instead, they are themselves the result of complicated simulations.

4.1 Four Roads from Matrix Elements to Answers

4.1.1 Analytic Integration

Analytic integration is mentioned only for completeness. Already for three or more particles in the final state with halfway realistic cuts, a fully analytical calculation of integrated cross sections is impossible. At least some integrations will have to be performed numerically.

4.1.2 Semianalytic Calculation

The radiative corrections for LEP1 precision physics on the Z^0-resonance are dominated by soft photons. The phase space in the soft limit remains sufficiently simple so that all but a few integrations can be performed analytically for simple cuts. Since programs based on this approach are extremely fast compared to the more universally applicable, they play an important rôle in LEP1 phenomenology. Due to their speed and precision, they are the preferred tool for the unfolding of electroweak precision observables at LEP1. An additional application is to act as benchmarks for other programs. The situation at LEP2 and the LC is different, because more phase space for hard radiation is available and substantial changes are necessary in these programs.

4.1.3 Monte Carlo Integration

If only $I(f)$ in (4.1) has to be calculated, it is possible to calculate the estimator $E(f)$, given in (4.7) below, directly, using an appropriate source of pseudo random numbers. This is known colloquially as "weighted event generation," because the phase space points used for sampling are not equally likely, but have a weight attached to them. Since f is not required to be bounded for this approach, the application can focus all attention on minimizing the sampling error (4.8). The techniques for achieving this goal are well developed and the currently most sophisticated are described in detail in section 4.3.

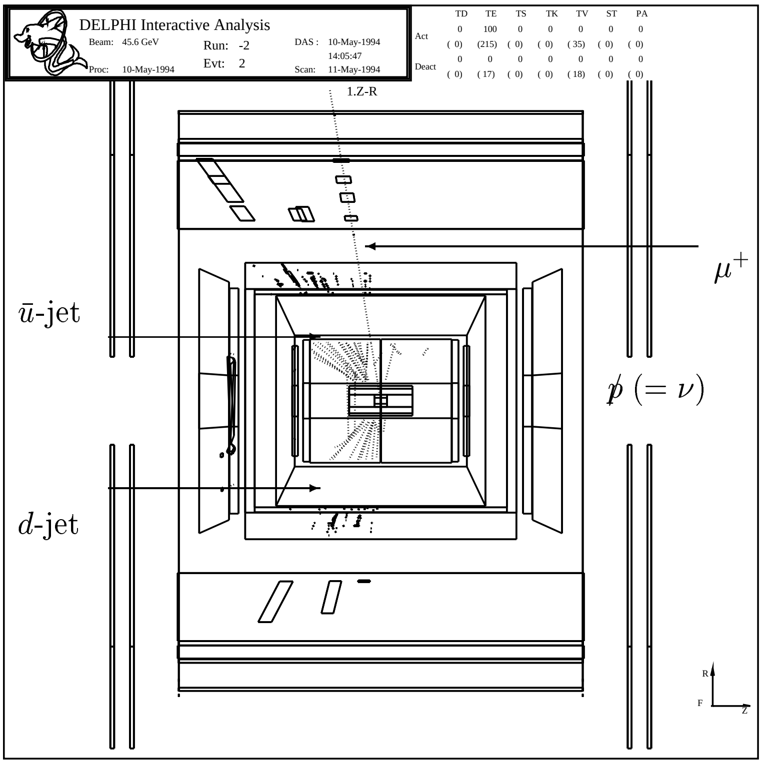


Figure 4.1: *Simulated event $e^+e^- \to d\bar{u}\mu\bar{\nu}_\mu$ at $\sqrt{s} = 200\,\mathrm{GeV}$ ("particle identification" from Monte Carlo record). Event generated by* **WOPPER** *[59] and passed through DELPHI detector simulation [60].*

4.1.4 Monte Carlo Event Generation

In principle, it is possible to use weighted event generators also in detector simulations similar to the one shown in figure 4.1, as long as the detector simulation is equipped with the trivial infrastructure for the bookkeeping of event weights.

However, this approach has the serious drawback that the distribution of events that are simulated in the detector, can be different from the realistic distributions. In the result of the integration, this is corrected by the weights, but the problem remains that the detector is exercised mostly in the region of the phase space where few events are expected in the experiment.

Therefore it is desirable to generated "unweighted events", i. e. events with uniform weight 1, which are distributed exactly as predicted by theory. For bounded probability distributions, such samples can be generated

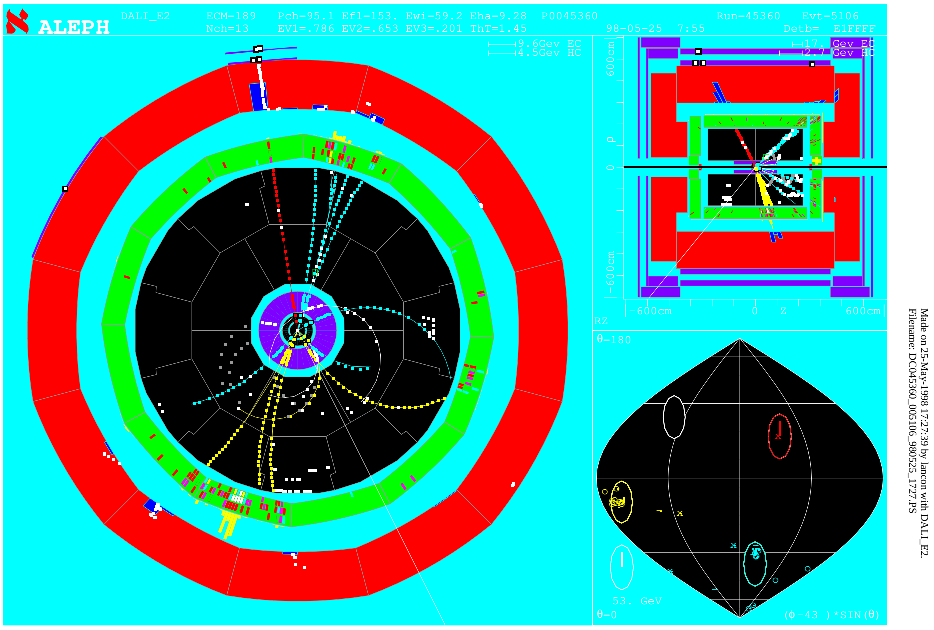


Figure 4.2: *Actual LEP2 event at* $\sqrt{s} = 189\,\mathrm{GeV}$*:* $e^+e^- \to q\bar{q}\mu\bar{\nu}_\mu$ *in the ALEPH detector.*

by brute force rejection (generate a phase space point x and a uniform deviate y; accept x, if $y \leq p(x)/p_{\max}$, else repeat) and if the distribution does not vary too much, this method can be sufficiently efficient. Unfortunately, most distributions in High Energy Physics (HEP) are singular and require sophisticated mapping techniques to handle the integrable singularities (see section 4.3).

4.1.5 Why We Can Not Use the Metropolis Algorithm Naively

At first sight, the Metropolis algorithm [61] appears to be an attractive alternative to the brute-force rejection algorithm for unweighted event generation. The Metropolis algorithm constructs a Markov process covering the phase space, which can be shown to converge asymptotically to the desired probability density without having to reject a single event. A Markov step is accepted with a probability that is equal to the ratio of the probability densities at the new and the old phase space point. If the move is rejected, the *same* point must be added to the sample *again*. The advantage of the procedure is that probabilities exceeding unity can be treated *as if* they were unity and probability densities need not be bounded for this agorithm to work (see [62, 63] for pedagogical discussions).

If the probability distribution is rather steep, simple numerical experiments show that the number of rejected moves is high. Therefore, the number of events that will appear more than once in a finite sample will be high as well and this causes problems further down the simulation chain. Error estimates in detector simulation and unfolding tools might not be prepared for finite samples with many duplicate events [64].

In general, the problem is that, unless the step size and the initial conditions are chosen (and varied!) carefully, errors can be *under*estimated significantly for steep distributions. Nevertheless, if the problems with obtaining *robust* error estimates can be gotten under control, the Metropolis algorithm could also develop into a powerful part of the calculational tool chest for event generators. But until then, a strong *caveat emptor* regarding results obtained in this way is in order.

4.2 Monte Carlo

simulate: *feign, pretend to have or feel, put on; pretend to be, act like, resemble, mimic, take or have an altered form suggested by; ...*
[CONCISE OXFORD DICTIONARY]

The overview in the preceding section has shown that weighted and unweighted Monte Carlo methods are the most robust and flexible part of the phenomenological tool chest for phase space integration. Therefore they deserve a more detailed discussion. Given a (pseudo-)random sequence of points in n-particle phase space

$$\mathfrak{p}_g = \{p_1, p_2, \ldots, p_N\}, \tag{4.4}$$

where the points are distributed according to

$$\mathrm{d}\mu_g(p) = g(p)\mathrm{d}\mu(p), \tag{4.5a}$$

with

$$\int_M \mathrm{d}\mu_g(p) = 1 \text{ and } g(p) > 0 \text{ (almost everywhere)}, \tag{4.5b}$$

we get an estimator of

$$I(f) = \int_M \mathrm{d}\mu(p)\, f(p) = \int_M \mathrm{d}\mu_g(p)\, \frac{f(p)}{g(p)} \tag{4.6}$$

as

$$E(f) = \left\langle \frac{f}{g} \right\rangle_g = \frac{1}{N} \sum_{i=1}^{N} \frac{f(p_i)}{g(p_i)} . \tag{4.7}$$

In (4.7), the volume of M is encoded in g, because g is normalized according to (4.5b). As written, $E(f)$ depends on g and the particular $\mathfrak{p}_g$, but the dependence on g drops out completely in the *ensemble average* over all $\mathfrak{p}_g$. Explicitely spelling out the dependence on g is important, because the measure $\mathrm{d}\mu(p)$ (i. e. $\mathrm{d}\mu_g(p)$ for $g(x) = 1$) is in general by no means the one that can be generated most efficiently. As will be shown shortly, it is also not the most useful one. Therefore a discussion should consider a general g from the outset.

For large $\mathfrak{p}_g$, $E(f)$ will converge to $I(f)$, but when comparing theoretical predictions with experimental observations, we have to deal with the more complicated problem of estimating the error for a finite sample size and minimizing this error.

Besides optimizing g for this purpose, we could also use a deterministic $\mathfrak{p}_g$ derived from *quasi random numbers*. Under certain conditions, this approach can speed up convergence dramatically [65] (from $1/\sqrt{N}$ to $1/N$) at the expense of the robustness of the conditions. Therefore we will not proceed along this line here and just notice that further research in the area is highly desirable. Also it is not clear how to interface deterministic Monte Carlos using quasi random numbers with detector simulations.

An unbiased estimator for the variance of $E(f)$ is given by

$$V(f,g) = \frac{1}{N-1} \left(\left\langle \left(\frac{f}{g} \right)^2 \right\rangle_g - \left\langle \frac{f}{g} \right\rangle_g^2 \right) . \tag{4.8}$$

A direct consequence of (4.8) is $V(f,f) = 0$. This vanishing error is not paradoxical, because $I(g)$ has to be known exactly in order to be able to generate $\mathfrak{p}_g$. For the following, it is useful to note that $V(f,g)$ is a biased estimator for $W_g(f/g)/(N-1)$, where the *variation* $W_g(h)$ of a function h with respect to the measure μ_g is defined as

$$W_g(h) = \int_M \mathrm{d}\mu_g(p) \, \left((h(p))^2 - (I(h))^2 \right) = \int_M \mathrm{d}\mu_g(p) \, \left(h(p) - I(h) \right)^2 \geq 0 \, . \tag{4.9}$$

The advantage of $W_g(f/g)$ over $V(f,g)$ is that $W_g(f/g)$ does *not* depend on the particular $\mathfrak{p}_g$, which will simplify the following reasoning considerably.

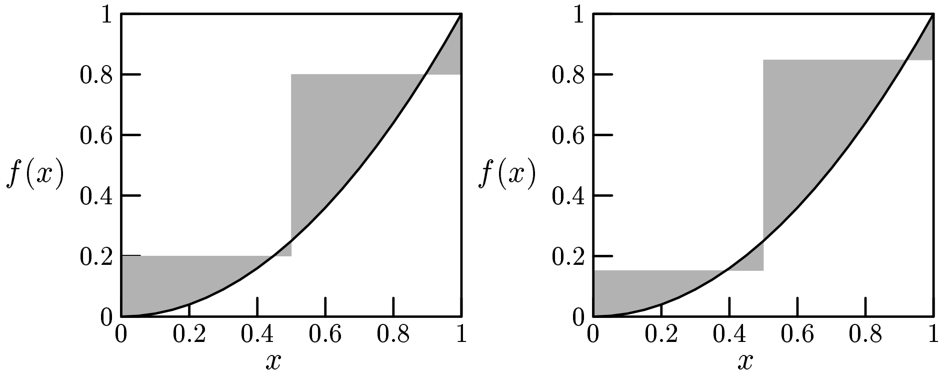


Figure 4.3: *In the left graph, $\Delta_1(f, g_\alpha)$ is minimized for a simple step function, while in the left graph $\Delta_2(f, g_\alpha)$ is minimized in the* same *function space. $\Delta_2(f, g_\alpha)$ distributes the mismatch more evenly to achieve faster convergence for integration, but has lower efficiency for event generation.*

For Monte Carlo integration, the optimal g will minimize

$$\Delta_2(f, g) = W_g(f/g) \tag{4.10}$$

while for event generation, the optimal g will minimize

$$\Delta_1(f, g) = \int_M \mathrm{d}\mu_g(p) \left(1 - \frac{\frac{f(p)}{g(p)}}{\sup_M \left(\frac{f}{g}\right)}\right) = 1 - \frac{I(f)}{\sup_M \left(\frac{f}{g}\right)}, \tag{4.11}$$

in order to maximize the rejection efficiency. If f is a wildly fluctuating function, this optimization of g is indispensable for obtaining a useful accuracy. Typical causes for large fluctuations are integrable singularities of f or μ inside of M or non-integrable singularities very close to M. Therefore, we will use the term "singularity" for those parts of M in which there are large fluctuations in f or μ.

If the g could be chosen from a function space that includes f, both Δ_1 and Δ_2 would find the same optimal $g = f$, since $\Delta_1(f, f) = 0$ and $\Delta_2(f, f) = 0$. However, in practical applications, we have $g \in G$, with a function space G that does *not* include f (otherwise we would already know both $I(f)$ and an efficient algorithm to generate phase space points according to μ_f). Therefore, which $g \in G$ is closest to f will in general depend on whether the metric is Δ_1 or Δ_2.

For concreteness, consider the trivial example of $f(x) = x^2$ on $M = [0,1]$ with a one parameter family of step functions

$$g_\alpha(x) = \alpha \cdot \Theta(1/2 - x) + (1-\alpha) \cdot \Theta(x - 1/2) \tag{4.12}$$

Then

$$\Delta_2(f, g_\alpha) = \frac{1}{160 \cdot \alpha} + \frac{155}{160 \cdot (1-\alpha)} - \frac{1}{9} \tag{4.13}$$

which is minimized by

$$\alpha_2 = \frac{1}{1+\sqrt{31}} = 0.15226\ldots. \tag{4.14}$$

On the other hand, $\Delta_1(f, g_\alpha)$ is minimized by

$$\alpha_1 = \frac{1}{5} \tag{4.15}$$

and there is a noticeable difference, even for smooth functions, as shown in figure 4.3.

From a technical point of view, Δ_2 is more convenient. For a given $g \in G$, $\Delta_2(f,g)$ will exist for many interesting functions f for which $\Delta_1(f,g)$ is useless because $\sup_M(f/g)$ diverges. Furthermore, the estimation of $\Delta_2(f,g)$ is statistically robust, while the estimation of $\sup_M(f/g)$ required by $\Delta_1(f,g)$ is much more delicate. Finally, the optimality equations for g derived from $\Delta_2(f,g)$ by functional derivation are linear.

The set G is not defined precisely, it includes all probability densities that can be generated "cheaply." This includes all densities for which the inverse of the anti-derivative can be computed efficiently, but there are other, less well known, cases [66]. Of course, there is no absolute measure for the "cheapness" of generating g. Instead, the effort has to be put in context. As long as the generation of a phase space point p does not take longer than the evaluation of the differential cross section $f(p)$, the generation can be considered "cheap." If the variance reduction is substantial, it can be useful to use even more expensive g's.

The manual minimization of either $\Delta_n(f,g)$ for a given f can be a tedious exercise, that has to be repeated every time the coupling parameters or kinematical cuts in f are changed. For this reason it is economical to use adaptive algorithms that minimize $\Delta_n(f,g)$ automatically, as long as they perform satisfactorily.

4.3 Adaptive Monte Carlo

Manual optimization of g is often too time consuming, in particular if the dependence of the integral on external parameters (in the integrand and in the boundaries) is to be studied. Adaptive numerical approaches are more attractive in these cases.

In this section, we will use for illustration only integrals over the unit hypercube

$$x \in M =]0,1[^{\otimes n} \tag{4.16}$$

with Cartesian measure

$$\mathrm{d}\mu(x) = \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \ldots \wedge \mathrm{d}x_n \tag{4.17}$$

and $\mathrm{Vol}(M) = 1$. The general case follows by covering the manifold M with one or more maps and taking into account the appropriate Jacobians.

4.3.1 Multi Channel

The basic question is: how can we implement a variable weight g in $\mathrm{d}\mu_g(x) = g(x)\mathrm{d}\mu(x)$ efficiently? An obvious option is to employ a weighted sum of densities

$$g(x) = \sum_{i=1}^{n} \alpha_i g_i(x) \tag{4.18}$$

with

$$\int \mathrm{d}\mu(x)\, g_i(x) = 1 \text{ and } \sum_{i=1}^{n} \alpha_i = 1 \text{ and } \alpha_i \geq 0 \tag{4.19}$$

and to optimize the weights α_i [67]. This approach has been very successful in Monte Carlo integration (e. g. [68, 69]). However, in applications with several hundred different channels g_i [69], the optimization of the weights α_i according to [67] yields only modest improvements [70]. It appears to be more important that a channel is present, than that the weights are accurately tuned.

In a typical application, the g_i are constructed manually from squared signal diagrams with simplified spin correlations. The corresponding distributions can be generated easily by successive integration of the s-channel poles and t-channel singularities. The drawback of this approach is that the location and the shape of the singularities has to be known analytically.

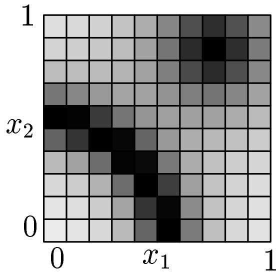


Figure 4.4: *A fixed grid with variable weights can not adapt to singular integrands.*

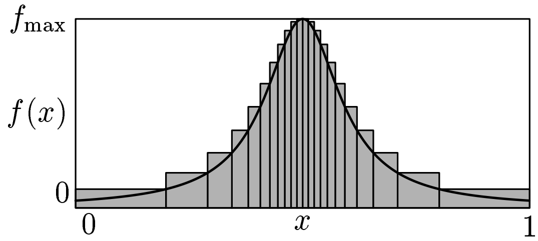


Figure 4.5: *In one dimension, a variable grid with fixed weights can adapt well to singular integrands.*

While the location of the poles is fixed by the particle masses, radiative correction and interference terms will modify the shape and shift the actual maxima. A particularly serious problem is that the effects of radiative corrections depend strongly on kinematical cuts.

A special case of (4.18) is given by using the characteristic functions of fixed bins as g_i, which corresponds to variable weights for bins in a fixed grid, as depicted in figure 4.4. This case is not useful for applications in particle physics, because a grid with a *fixed resolution* can not adapt to the power law singularities from propagators as are ubiquitous in particle physics. Differential cross sections typically vary over many orders of magnitude and the grid would have to be too fine to be useful.

In *one* dimension, there is an interesting alternative to varying the weights

Figure 4.6: *Quadrangles do not remain convex in general after unconstrained shifts of lattice points.*

of fixed width. It is possible to vary instead the size of bins with fixed weight[1]

$$g(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{\Theta(x - x_{i-1})\Theta(x_i - x)}{x_i - x_{i-1}} . \tag{4.20}$$

This way it is possible to have bins corresponding to densities that vary over many orders of magnitude. It is possible to iteratively improve the grid, driven by samplings of $f(x)\mathrm{d}\mu_g(x)$. The improvement can either decrease the distance $\Delta_1(f, g)$ (*importance sampling* for event generation) or decrease the distance $\Delta_2(f, g)$ (*stratified sampling* for high precision integration). An optimal g would solve the functional equation

$$\frac{\delta}{\delta g} \Delta_{1,2}(f, g) = \Delta'_{1,2}(f, g) = 0. \tag{4.21}$$

While it is straightforward to sample $\Delta'_{1,2}(f, g)$ simultaneously with f/g, an exact solution of (4.21) is not possible, because contributions have to be moved from one bin to another. Nevertheless, an approximate solution is possible under the assumption that the integrand in (4.21) is piecewise constant [71].

Unfortunately, there is no immediate generalization to higher dimensions, because freely shifting vertices starting from a hypercubic lattice will in general lead to quadrangles that are no longer convex, as illustrated in figure 4.6. The concave quadrangles are more than a technical inconvenience, because they lead to numeric instabilities of the grid optimization.

4.3.2 Delaunay Triangulation

Mathematically, the most natural decomposition of the integration region in higher dimension is into simplices. In fact, there have been attempts

[1] Varying the weights as well would be possible but adds no new flexibility. It is therefore better to avoid the added complexity and to keep all weights identical.

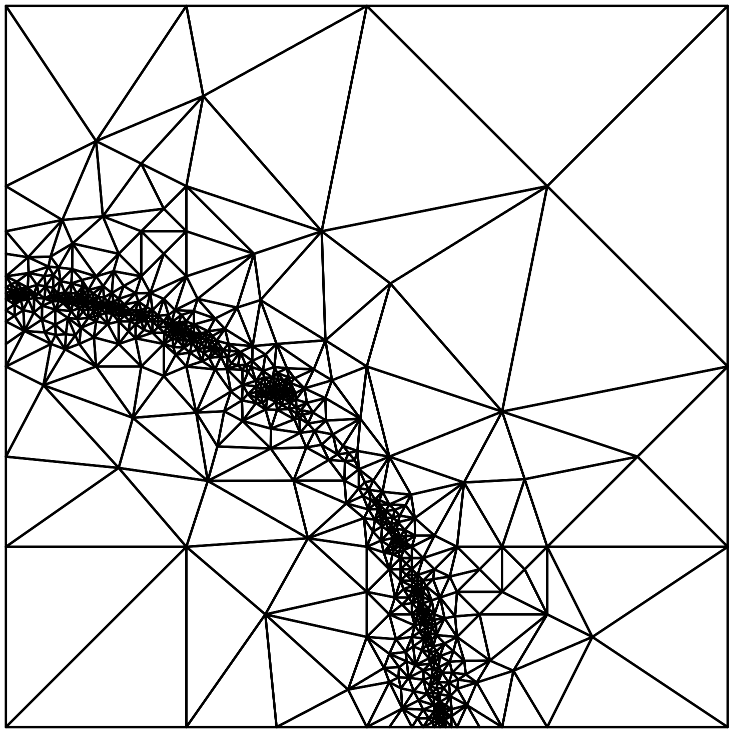


Figure 4.7: *Delaunay triangulation with Steiner points adapted to a singularity on a quarter circle.*

to use simplical decompositions in adaptive Monte Carlo [72]. The results of this exercise have not been impressive so far, in particular for strongly fluctuating integrands. The reason for the failure of the naive approach is that the shape of the simplices is not constrained and that instabilities will produce "splinters" that are numerically ill-behaved.

The construction of "well behaved" simplical decompositions is a rather non-trivial mathematical problem (see [73] for reviews), that remains unsolved in the general case. However, there has been progress in two and three dimensions, driven in part by the needs of the engineering community for numerically robust finite element codes. The theory for two dimensions is rather complete and mathematically very beautiful [74]. Therefore it is appropriate to make a factorized ansatz [75], in which the "least factorizable" pairs and triplets are handled by simplical decomposition, using triangles and tetrahedra.

In two dimensions, the most suitable tool are *constrained Delaunay triangulations* with *Steiner points*. For a given set of points, the Delaunay triangulation is the unique triangulation in which the center of the circle passing through the corners of each triangle is inside of that triangle. This uniqueness helps to stabilize the optimization of the triangulation, which is no longer unique, if we allow to add additional points, so-called Steiner points, to the point set in order to satisfy additional constraints. Nevertheless, there are well defined and terminating algorithms [76] that construct constrained Delaunay triangulations with global and local maximum area constraints, as well as minimum angle constraints. The former are what is required for modeling a function for importance sampling and the latter stabilize the iterative refinement of the triangulation.

However, there is still much more work needed to put these ideas to work in particle physics applications beyond a proof of concept [75], because the required computational geometry is at or beyond the state-of-the-art in finite element analysis for engineering.

4.3.3 VEGAS

Factorization

The previous section has shown that one-dimensional distributions are singled out because a mathematically well defined optimization strategy can be formulated with very little effort. Therefore it was tempting to propose a factorized ansatz [71]

$$g(x) = g_1(x_1) g_2(x_2) \cdots g_n(x_n) \tag{4.22}$$

and to investigate its properties. It turns out that (4.22) produces satisfactory results with very little effort and the corresponding code `VEGAS` [71] has been used as a “black box” by countless physics programs. For two decades no significant improvements of `VEGAS` have been proposed.

The factorized ansatz (4.22) of `VEGAS` guarantees that the resulting grid is hypercubic. Furthermore, the simple one-dimensional update prescriptions can be used in each dimension, if the estimators entering the prescriptions are averaged over the remaining dimensions. A final benefit of the factorized ansatz is that the computational costs grow only linearly with the number of dimensions and not exponentially.

Of course, not all functions of interest in particle physics can be approximated by a factorized ansatz. While the functions depicted in figure 4.8 have factorized singularities after a simple coordinate transformation, their sum in figure 4.9 can not be sampled efficiently by `VEGAS`. Similar example are ubiquitous in particle physics, because the cross sections are derived from sums

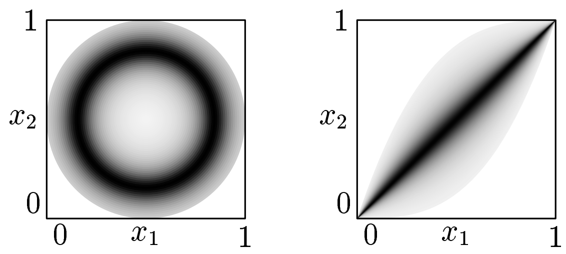


Figure 4.8: *Integrands that **VEGAS** can handle after a coordinate transformation.*

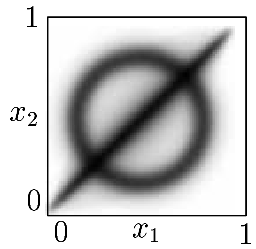


Figure 4.9: *An integrand that **VEGAS** can not handle, even after a coordinate transformation.*

of Feynman diagrams and have singularities in several kinematical variables which can not be used as coordinates simultaneously.

For example, the three jet production $e^+e^- \to q\bar{q}g$ in figure 4.10 has important singularities in both variables $s_{1/2} = (q_{1/2}+k)^2$. In this case, there are parametrizations of the phase space that use both s_1 and s_2 simultaneously. However the corresponding Jacobian $|\partial\phi/\partial x|$ contains non-factorizable singularities in s_1 and s_2 induced by a Gram determinant as $1/\sqrt{\Delta_4(p_1,p_2,q_1,q_2)}$. Thus there is no parametrizations in which both singularities factorize simultaneously and `VEGAS` can not adjust to both singularities (see [77]). A general solution to this problem will be presented below in section 4.3.4.

Stratified vs. Importance Sampling

The implementation of `VEGAS` [71] that is widely used, has three different modes: importance sampling, stratified sampling and a hybrid of the two. Importance sampling can be selected by the application program, while the choice between stratified sampling and the hybrid is done by `VEGAS` based

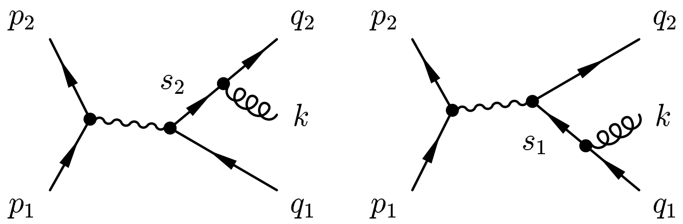


Figure 4.10: *Three jet production $e^+e^- \to q\bar{q}g$.*

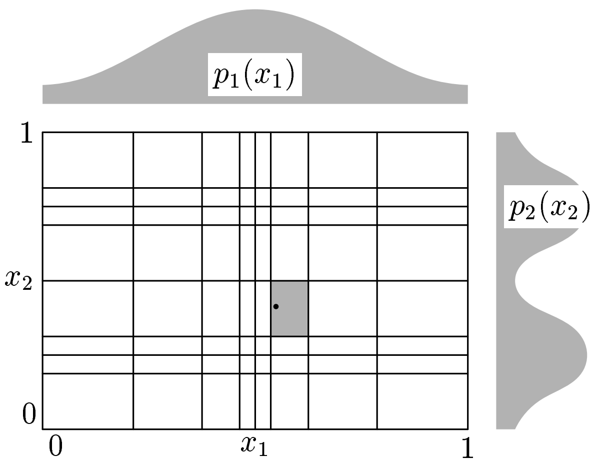


Figure 4.11: ***VEGAS** grid structure for importance sampling*

on the dimension of the integration domain. The implementation [71] very terse and hard to understand. Even popular textbooks [78] reproduce the code without further explanation. The reimplementation that was required for [79], aims to be more readable.

In pure importance sampling, `VEGAS` adjusts a grid like the two-dimensional illustration in figure 4.11. The sampling proceeds by first selecting a hypercube at random and subsequently selecting a point in this hypercube according to a uniform distribution. This algorithm generates points distributed according to (4.20) and (4.22).

In stratified sampling, faster convergence is attempted by distributing sampling points more uniformly than random samples. In `VEGAS`, stratified sampling is implemented by dividing the bins of the adaptive grid further as shown in figure 4.12. The number of divisions is chosen such that the sum of two samples per bin approximates the desired total number of samples.

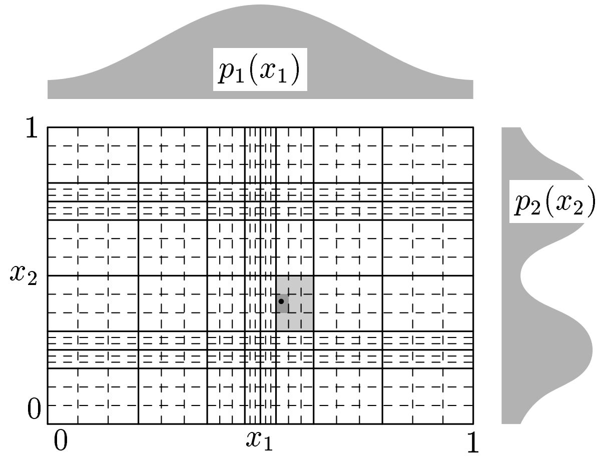


Figure 4.12: ***VEGAS** grid structure for genuinely stratified sampling.*

Now the sampling proceeds by selecting two random points from *each* bin of the finer grid. This procedure shows better convergence for integration than importance sampling.

As can be seen from table 4.1, the algorithm for stratified sampling as described above can only be used in low dimensions. Already the eight-dimensional four-particle phase space would require many billions of sampling points, which is usually not possible.

Nevertheless, the stratified sampling algorithm performs empirically better than the importance sampling algorithm and it is worthwhile to pursue the hybrid approach illustrated in figure 4.13. Here the stratification grid is no longer a subdivision of the adaptive grid. Instead, the points selected in

Table 4.1: *To stratify or not to stratify: number of calls required for a grid of 25 bins in each direction, as a function of the dimension of the integration domain.*

n_{dim}	$N_{\text{calls}}^{\max}(n_g = 25)$
2	$1 \cdot 10^3$
3	$3 \cdot 10^4$
4	$8 \cdot 10^5$
5	$2 \cdot 10^7$
6	$5 \cdot 10^8$

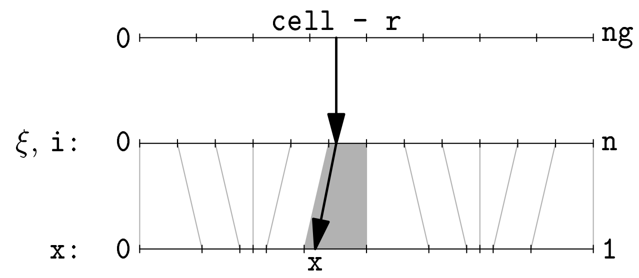


Figure 4.13: *VEGAS grid structure for pseudo stratified sampling.*

the stratification grid are mapped linearly onto the adaptive grid and further nonlinearly onto the integration domain. The adaptive grid is updated to minimize $\Delta_1(f,g)$. This hybrid is, strictly speaking, theoretically not justified because importance sampling assumes a *random* distribution. However, the "more uniform" distribution appears to perform better in practical applications and the hybrid approach is very successful.

4.3.4 VAMP

Introduction

As mentioned before for the examples in figure 4.8, the property of factorization depends on the coordinate system. Consider, for example, the functions

$$f_1(x_1,x_2) = \frac{1}{(x_1-a_1)^2+b_1^2} \tag{4.23a}$$

$$f_2(x_1,x_2) = \frac{1}{\left(\sqrt{x_1^2+x_2^2}-a_2\right)^2+b_2^2} \tag{4.23b}$$

on $M = (-1,1)\otimes(-1,1)$ with the measure $\mathrm{d}\mu = \mathrm{d}x_1 \wedge \mathrm{d}x_2$. Obviously, f_1 is factorizable in Cartesian coordinates, while f_2 is factorizable in polar coordinates. `VEGAS` will sample either function efficiently for arbitrary $b_{1,2}$ in suitable coordinate systems, but there is no coordinate system in which `VEGAS` can sample the sum f_1+f_2 efficiently for small $b_{1,2}$.

There is however a generalization [80] of the `VEGAS` algorithm from factorizable distributions to sums of factorizable distributions, where each term may be factorizable in a *different* coordinate system. This larger class includes most of the integrands appearing in particle physics and empirical studies have shown a dramatic increase of accuracy for typical integrals.

The problem of estimating $I(f)$ can be divided naturally into two parts: parametrization of M and sampling of the function f. While the estimate will not depend on the parametrization, the error will.

In general, we need an atlas with more that one chart ϕ to cover a manifold M. We will ignore this technical complication in the following, because, for the purpose of integration, we can always decompose M such that each piece is covered by a single chart. Moreover, a single chart suffices in most cases of practical interest, since we are at liberty to remove sets of measure zero from M. For example, after removing a single point, the unit sphere can be covered by a single chart.

Nevertheless, even if we are not concerned with the global properties of M that require the use of more than one chart, the language of differential geometry will allow us to use our geometrical intuition. Instead of pasting together locally flat pieces, we will paste together *factorizable* pieces, which can be overlapping, because integration is an additive operation.

For actual computations, it is convenient to use the same domain for the charts of all manifolds. The obvious choice for n-dimensional manifolds is the open n-dimensional unit hypercube

$$U = (0,1)^{\otimes n}\,. \tag{4.24}$$

Sometimes, it will be instructive to view the chart as a composition $\phi = \psi \circ \chi$ with an irregularly shaped $P \in \mathbf{R}^n$ as an intermediate step

$$\begin{tikzcd} & U \arrow[ddl, "\chi"'] \arrow[d, "\phi"] \arrow[ddr, dotted, "f\circ\phi"] & \\ & M \arrow[dr, dotted, "f"] & \\ P \arrow[ur, "\psi"] \arrow[rr, dotted, "f\circ\psi"'] & & \mathbf{R} \end{tikzcd} \tag{4.25}$$

(in all commutative diagrams, solid arrows are reserved for bijections and dotted arrows are used for other morphisms). The integral (4.1) can now be written

$$I(f) = \int_0^1 \mathrm{d}^n x \left|\frac{\partial\phi}{\partial x}\right| f(\phi(x)) \tag{4.26}$$

and it remains to sample $|\partial\phi/\partial x| \cdot (f \circ \phi)$ on U. Below, it will be crucial that there is always more than one way to map U onto M

$$\begin{tikzcd}
U \arrow[rr, "\chi"] \arrow[dr, "\phi"] \arrow[dd, "\pi_U"'] & & P \arrow[dl, "\psi"'] \arrow[dd, "\pi_P"] \\
& M & \\
U' \arrow[rr, "\chi'"] \arrow[ur, "\phi'"] & & P' \arrow[ul, "\psi'"']
\end{tikzcd} \tag{4.27}$$

and that we are free to select the map most suitable for our purposes.

The ideal choice for ϕ would be a solution of the partial differential equation $|\partial\phi/\partial x| = 1/(f \circ \phi)$, but this is equivalent to an analytical evaluation of $I(f)$ and is impossible for the cases under consideration. A more realistic goal is to find a ϕ such that $|\partial\phi/\partial x| \cdot (f \circ \phi)$ has factorizable singularities and is therefore sampled well by `VEGAS`. This is still a non-trivial problem, however. For example, consider again the phase space integration for gluon radiation $e^+e^- \to q\bar{q}g$. From the Feynman diagrams in figure 4.10 we had seen that the squared matrix element has singularities in both variables $s_{1/2} = (q_{1/2} + k)^2$ that can not be factorized. On the other hand, it is straightforward to find parametrizations that factorize the dependency on s_1 or s_2 *separately*.

Returning to the general case, consider N_c different maps $\phi_i : U \to M$ and probability densities $g_i : U \to [0, \infty)$. Then the function

$$g = \sum_{i=1}^{N_c} \alpha_i (g_i \circ \phi_i^{-1}) \left| \frac{\partial \phi_i^{-1}}{\partial p} \right| \tag{4.28}$$

is a probability density $g : M \to [0, \infty)$

$$\int_M \mathrm{d}\mu(p)\, g(p) = 1\,, \tag{4.29}$$

as long as the g_i are probability densities themselves and the α_i are properly normalized

$$\int_0^1 g_i(x) \mathrm{d}^n x = 1\,, \quad \sum_{i=1}^{N_c} \alpha_i = 1\,, \quad 0 \le \alpha_i \le 1\,. \tag{4.30}$$

From the definition (4.28), we have obviously

$$I(f) = \sum_{i=1}^{N_c} \alpha_i \int_M g_i(\phi_i^{-1}(p)) \left| \frac{\partial \phi_i^{-1}}{\partial p} \right| \mathrm{d}\mu(p)\, \frac{f(p)}{g(p)} \tag{4.31}$$

and, after pulling back from M to N_c copies of U

$$I(f) = \sum_{i=1}^{N_c} \alpha_i \int_0^1 g_i(x) \mathrm{d}^n x \, \frac{f(\phi_i(x))}{g(\phi_i(x))} \,, \tag{4.32}$$

we find a new estimator of the integral $I(f)$

$$E(f) = \sum_{i=1}^{N_c} \alpha_i \left\langle \frac{f \circ \phi_i}{g \circ \phi_i} \right\rangle_{g_i} . \tag{4.33}$$

If the g_i in (4.31) and (4.33) are factorized, they can be optimized using the classic `VEGAS` algorithm [71] unchanged. However, since we have to sample with a separate adaptive grid for each channel, a new implementation [79] is required for technical reasons.

Using the N_c^2 maps $\pi_{ij} = \phi_j^{-1} \circ \phi_i : U \to U$ introduced in (4.27), we can write the $g \circ \phi_i : U \to [0, \infty)$ in (4.33) as

$$g \circ \phi_i = \left| \frac{\partial \phi_i}{\partial x} \right|^{-1} \left(\alpha_i g_i + \sum_{\substack{j=1 \\ j \neq i}}^{N_c} \alpha_j (g_j \circ \pi_{ij}) \left| \frac{\partial \pi_{ij}}{\partial x} \right| \right) . \tag{4.34}$$

From a geometrical perspective, the maps π_{ij} are just the coordinate transformations from the coordinate systems in which the other singularities factorize into the coordinate system in which the current singularity factorizes.

Even if all g_i factorize, the $g \circ \phi_i$ will not necessarily factorize, provided the ϕ_i are chosen appropriately. The maps π_{ij} are purely geometric objects, independent of the physical model under consideration. They are the coordinate transformation from a coordinate system Ξ_i in which the ith singularity factorizes to a coordinate system Ξ_j in which the jth singularity factorizes.

The construction of the π_{ij} requires some mathematical input, but it is required only once and the procedure allows economical studies of the dependence of the cross section on kinematical cuts and on external parameters through f, because `VEGAS` can optimize the g_i for each parameter set and set of cuts without further human intervention.

Note that the integral in (4.31) does not change, when we use $\phi_i : U \to M_i \supseteq M$, if we extent f from M to M_i by the definition $f(M_i \setminus M) = 0$. This is useful, for instance, when we want to cover $(-1,1) \otimes (-1,1)$ by both Cartesian and polar coordinates. This causes, however, a problem with the π_{12} in (4.34). In the diagram

$$
\begin{tikzcd}
P_1 \arrow[r, "\psi_1"] & M_1 & M \arrow[l, dotted, "\iota_1"'] \arrow[r, dotted, "\iota_2"] & M_2 & P_2 \arrow[l, "\psi_2"'] \\
& U \arrow[ul, "\chi_1"] \arrow[u, "\phi_1"'] \arrow[rr, dotted, "\pi_{12}"] & & U \arrow[u, "\phi_2"] \arrow[ur, "\chi_2"'] &
\end{tikzcd}
\tag{4.35}
$$

the injections $\iota_{1,2}$ are not onto and since π_{12} is not necessarily a bijection anymore, the Jacobian $|\partial\pi_{ij}/\partial x|$ may be ill-defined. But since $f(M_i \backslash M) = 0$, we only need the unique bijections $\phi'_{1,2}$ and π'_{12} that make the diagram

$$
\begin{tikzcd}
P_1 \arrow[r, "\psi_1"] & M_1 & M \arrow[l, dotted, "\iota_1"'] \arrow[r, equal] & M \arrow[r, dotted, "\iota_2"] & M_2 & P_2 \arrow[l, "\psi_2"'] \\
& U \arrow[ul, "\chi_1"] \arrow[u, "\phi_1"'] & U_1 \arrow[l, dotted, "\iota_1^U"] \arrow[u, "\phi'_1"] \arrow[r, "\pi'_{12}"] & U_2 \arrow[u, "\phi'_2"'] \arrow[r, dotted, "\iota_2^U"'] & U \arrow[u, "\phi_2"] \arrow[ur, "\chi_2"'] &
\end{tikzcd}
\tag{4.36}
$$

commute.

Multi Channel Sampling

Up to now, we have not specified the α_i, they are only subject to the conditions (4.30). Intuitively, we expect the best results when the α_i are proportional to the contribution of their corresponding singularities to the integral. The option of tuning the α_i manually is not attractive if the optimal values depend on varying external parameters. Instead, we use a numerical procedure [67] for tuning the α_i.

We want to minimize the variance (4.8) with respect to the α_i. This is equivalent to minimizing

$$
W(f,\alpha) = \int_M g(p)\mathrm{d}\mu(p) \left(\frac{f(p)}{g(p)}\right)^2 \tag{4.37}
$$

with respect to α with the subsidiary condition $\sum_i \alpha_i = 1$. After adding a Lagrange multiplier, the stationary points of the variation are given by the solutions to the equations

$$
\forall i : W_i(f,\alpha) = W(f,\alpha) \tag{4.38}
$$

where

$$
W_i(f,\alpha) = -\frac{\partial}{\partial\alpha_i} W(f,\alpha) = \int_0^1 g_i(x)\mathrm{d}^n x \left(\frac{f(\phi_i(x))}{g(\phi_i(x))}\right)^2 \tag{4.39}
$$

and

$$W(f,\alpha) = \sum_{i=1}^{N_c} \alpha_i W_i(f,\alpha)\,. \tag{4.40}$$

It can easily be shown [67] that the stationary points (4.38) correspond to local minima. If we use

$$N_i = \alpha_i N \tag{4.41}$$

to distribute N sampling points among the channels, the $W_i(f,\alpha)$ are just the contributions from channel i to the total variance. Thus we recover the familiar result from stratified sampling, that the overall variance is minimized by spreading the variance evenly among channels.

The $W_i(f,\alpha)$ can be estimated with very little extra effort while sampling $I(f)$ (see (4.33))

$$V_i(f,\alpha) = \left\langle \left(\frac{f \circ \phi_i}{g \circ \phi_i} \right)^2 \right\rangle_{g_i} \,. \tag{4.42}$$

Note that the factor of g_i/g in the corresponding formula in [67] is absent from (4.42), because we are already sampling with the weight g_i in each channel separately.

The equations (4.38) are non linear and can not be solved directly. However, the solutions of (4.38) are a fixed point of the prescription

$$\alpha_i \mapsto \alpha_i' = \frac{\alpha_i \left(V_i(f,\alpha)\right)^\beta}{\sum_i \alpha_i \left(V_i(f,\alpha)\right)^\beta}\,, \quad (\beta > 0) \tag{4.43}$$

for updating the weights α_i. There is no guarantee that this fixed point will be reached from a particular starting value, such as $\alpha_i = 1/N_c$, through successive applications of (4.43). Nevertheless, it is clear that (4.43) will concentrate on the channels with large contributions to the variance, as suggested by stratified sampling. Furthermore, empirical studies show that (4.43) is successful in practical applications. The value $\beta = 1/2$ has been proposed in [67], but it can be beneficial in some cases to use smaller values like $\beta = 1/4$ to dampen statistical fluctuations.

Performance

Both the implementation and the practical use of the multi channel algorithm are more involved than the application of the original `VEGAS` algorithm.

Therefore it is necessary to investigate whether the additional effort pays off in terms of better performance.

A Fortran implementation of this algorithm, `VAMP` [79], has been used for empirical studies. This implementation features other improvements over "`VEGAS` Classic", most notably system independent and portable support for parallel processing (see section 4.4) and support for unweighted event generation.

There are two main sources of additional computational costs: at each sampling point the function $g \circ \phi_i$ has be evaluated, which requires the computation of the $N_c - 1$ maps π_{ij} together with their Jacobians and of the $N_c - 1$ probability distributions g_i of the other `VEGAS` grids (see (4.34)).

The retrieval of the current g_i's requires a bisection search in each dimension, i.e. a total of $O((N_c - 1) \cdot n_{\text{dim}} \cdot \log_2(n_{\text{div}}))$ executions of the inner loop of the search. For simple integrands, this can indeed be a few times more costly than the evaluation of the integrand itself.

The computation of the π_{ij} can be costly as well. However, unlike the g_i, this computation can usually be tuned manually. This can be worth the effort if many estimations of similar integrals are to be performed. Empirically, straightforward implementations of the π_{ij} add costs of the same order as the evaluation of the g_i.

Finally, additional iterations are needed for adapting the weights α_i of the multi channel algorithm described in (4.3.4). Their cost is negligible, however, because they are usually performed with far fewer sampling points than the final iterations.

Even in cases in which the evaluation of g_i increases computation costs by a whole order of magnitude, any reduction of the error by more than a factor of 4 will make the multi channel algorithm economical. In fact, it is easy to construct examples in which the error will be reduced by more than two orders of magnitude. The function

$$f(x) = \frac{b}{144\,\text{atan}(1/2b)} \left(\frac{3\pi\Theta(r_3 < 1)}{r_3^2((r_3 - 1/2)^2 + b^2)} + \frac{2\pi\Theta(r_2 < 1, |x_3| < 1)}{r_2((r_2 - 1/2)^2 + b^2)} + \frac{\Theta(-1 < x_1, x_2, x_3 < 1)}{x_1^2 + b^2} \right), \quad (4.44)$$

with $r_2 = \sqrt{x_1^2 + x_2^2}$ and $r_3 = \sqrt{x_1^2 + x_2^2 + x_3^2}$, is constructed such that it can easily be normalized

$$\int_{-1}^{1} \mathrm{d}^3 x\, f(x) = 1 \quad (4.45)$$

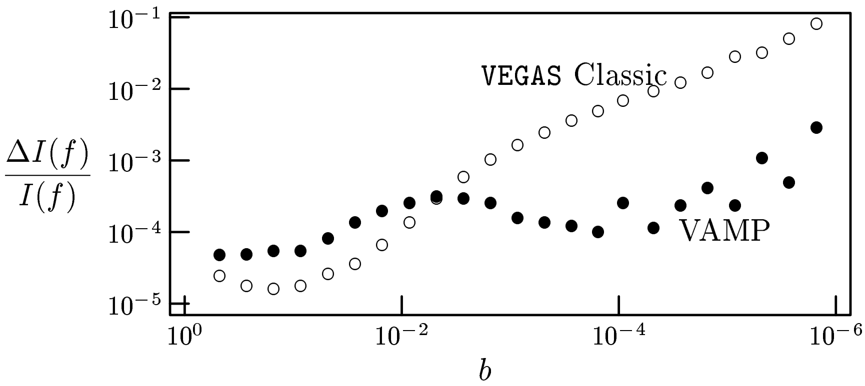


Figure 4.14: *Comparison of the sampling error for the integral of f in (4.44) as a function of the width parameter b for the two algorithms at comparable computational costs.*

and allows a check of the result. The three terms factorize in spherical, cylindrical and Cartesian coordinates, respectively, suggesting a three channel approach. After five steps of weight optimization consisting four iterations of 10^5 samples, we have performed three iterations of 10^6 samples with the VAMP multi channel algorithm. Empirically, we found that we can perform four iterations of $5 \cdot 10^5$ samples and three iterations of $5 \cdot 10^6$ samples with the classic `VEGAS` algorithm during the same time period. Since the functional form of f is almost as simple as the coordinate transformation, the fivefold increase of computational cost is hardly surprising.

In figure 4.14, we compare the error estimates derived by the classic `VEGAS` algorithm and by the three channel VAMP algorithm. As one would expect, the multi channel algorithm does not offer any substantial advantages for smooth functions (i. e. $b > 0.01$). Instead, it is penalized by the higher computational costs. On the other hand, the accuracy of the classic `VEGAS` algorithm deteriorates like a power with smaller values of b. At the same time, the multi channel algorithm can adapt itself to the steeper functions, leading to a much slower loss of precision.

The function f in (4.44) has been constructed as a showcase for the multi channel algorithm, of course. Nevertheless, more complicated realistic examples from particle physics appear to gain about an order of magnitude in accuracy. Furthermore, the new algorithm allows *unweighted* event generation. This is hardly ever possible with the original `VEGAS` implementation, because the remaining fluctuations typically reduce the average weight to very small numbers.

A particularly attractive application is provided by automated tools for the calculation of scattering cross sections. While these tools can currently calculate differential cross sections without manual intervention, the phase space integrations still require hand tuning of mappings for importance sampling for each parameter set. The present algorithm can overcome this problem, since it requires to solve the geometrical problem of calculating the maps π_{ij} in (4.34) for all possible invariants only *once*. The selection and optimization of the channels can then be performed algorithmically.

A Cheaper Alternative

There is an alternative approach that avoids the evaluation of the g_i's, sacrificing flexibility. Fixing the g_i at unity, we have for $\tilde{g} : M \to [0, \infty)$

$$\tilde{g} = \sum_{i=1}^{N_c} \alpha_i \left| \frac{\partial \phi_i^{-1}}{\partial p} \right| \tag{4.46}$$

and the integral becomes

$$I(f) = \sum_{i=1}^{N_c} \alpha_i \int_M \left| \frac{\partial \phi_i^{-1}}{\partial p} \right| \mathrm{d}\mu(p) \, \frac{f(p)}{\tilde{g}(p)} = \sum_{i=1}^{N_c} \alpha_i \int_0^1 \mathrm{d}^n x \, \frac{f(\phi_i(x))}{\tilde{g}(\phi_i(x))} \,. \tag{4.47}$$

`VEGAS` can now be used to perform adaptive integrations of

$$I_i(f) = \int_0^1 \mathrm{d}^n x \, \frac{f(\phi_i(x))}{\tilde{g}(\phi_i(x))} \tag{4.48}$$

individually. In some cases it is possible to construct a set of ϕ_i such that $I_i(f)$ can estimated efficiently. The optimization of the weights α_i can again be effected by the multi channel algorithm described in (4.3.4).

The disadvantage of this approach is that the optimal ϕ_i will depend sensitively on external parameters and the integration limits. In the approach based on the g in (4.28), `VEGAS` can take care of the integration limits automatically.

4.4 Parallelization

The computing needs of the LC can only be satisfied economically if parallel computing on commodity hardware can be realized. Traditionally, parallelization is utilized in HEP by running independent Monte Carlo or analysis jobs in parallel and combining their results statistically. Nevertheless, more fine grained parallelism is desirable in Monte Carlo jobs that use adaptive algorithms[2].

The problem of the parallelization of adaptive Monte Carlo integration algorithms, in particular `VEGAS`, has gained some attention recently [81, 82]. These implementations start from the classic implementation of `VEGAS` and add synchronization barriers, either mutexes for threads accessing shared memory or explicit message passing. These approaches result in compact code and achieve high performance, but their low level nature obscures the mathematical structure. Since parallel computing is notoriously difficult, it is desirable to have a well defined mathematical model that frees physicists from having to waste too much effort dealing with technical computing details. Therefore, we suggest a *mathematical* model of parallelism for adaptive Monte Carlo integration that is independent both of a concrete paradigm for parallelism and of the programming language used for an implementation. We decompose the algorithm and prove that certain parts can be executed in *any* order without changing the result. As a corollary, we know that they can be executed in parallel.

The algorithms presented below have been implemented successfully in the library `VAMP` [79], along with the multi channel algorithm described in section 4.3.4.

4.4.1 Vegas

As described in section 4.3.3, `VEGAS` uses two different grids: an adaptive grid G^A, which is used to adapt the distribution of the sampling points and a stratification grid G^S for stratified sampling. The latter is static and depends only on the number of dimensions and on the number of sampling points. Both grids factorize into *divisions* $d^i_{A,S}$

$$G^A = d_1^A \otimes d_2^A \otimes \cdots \otimes d_n^A \tag{4.49a}$$

$$G^S = d_1^S \otimes d_2^S \otimes \cdots \otimes d_n^S \,. \tag{4.49b}$$

The divisions come in three kinds

$$d_i^S = \emptyset \qquad \text{(importance sampling)} \tag{4.50a}$$

$$d_i^A = d_i^S / m \qquad \text{(stratified sampling)} \tag{4.50b}$$

$$d_i^A \neq d_i^S / m \qquad \text{(pseudo-stratified sampling)} \,. \tag{4.50c}$$

[2]Already at LEP2, there are problems with the parallelization of unweighted event generation if the maximum weights have to be found in each parallel job separately [64].

In the classic implementation of `VEGAS` [71], all divisions are of the same type. In a more general implementation [79], this is not required and it can be useful to use stratification only in a few dimensions.

Two-dimensional grids for the cases (4.50a) and (4.50b) have been illustrated in figures 4.11 and 4.12. In case (4.50a), there is no stratification grid and the points are picked at random in the whole region according to G_A. In case (4.50b), the adaptive grid G_A is a regular subgrid of the stratification grid G_S and an equal number of points are picked at random in each cell of G_S. Since $d_i^A = d_i^S/m$, the points will be distributed according to G_A as well. A one-dimensional illustration of the most complicated case (4.50c) has been shown in figure 4.13.

4.4.2 Formalization of Adaptive Sampling

In order to discuss the problems with parallelizing adaptive integration algorithms and to present a solution, it helps to introduce some mathematical notation. A sampling S is a map from the space π of point sets and the space F of functions to the real (or complex) numbers

$$\begin{aligned} S : \pi \times F &\to \mathbf{R} \\ (p, f) &\mapsto I = S(p, f)\,. \end{aligned} \tag{4.51}$$

For our purposes, we have to be more specific about the nature of the point set $p \in \pi$. In general, this point set will be characterized by a sequence of pseudo random numbers $\rho \in R$ and by one or more grids $G \in \Gamma$ used for importance or stratified sampling. A simple sampling

$$\begin{aligned} S_0 : R \times \Gamma \times A \times F \times \mathbf{R} \times \mathbf{R} &\to R \times \Gamma \times A \times F \times \mathbf{R} \times \mathbf{R} \\ (\rho, G, a, f, \mu_1, \mu_2) &\mapsto (\rho', G, a', f, \mu_1', \mu_2') = S_0(\rho, G, a, f, \mu_1, \mu_2) \end{aligned} \tag{4.52}$$

estimates the n-th moments $\mu_n' \in \mathbf{R}$ of the function $f \in F$. The integral and its standard deviation can be derived easily from the moments

$$I = \mu_1 \tag{4.53a}$$

$$\sigma^2 = \frac{1}{N-1}\left(\mu_2 - \mu_1^2\right)\,, \tag{4.53b}$$

while the latter are more convenient for the following discussion. In addition, S_0 collects auxiliary information to be used in the grid refinement, denoted by $a \in A$. The unchanged arguments G and f have been added to the result of S_0 in (4.52), so that S_0 has identical domain and codomain and can therefore be iterated. Previous estimates μ_n may be used in the estimation of μ'_n, but a particular S_0 is free to ignore them as well. Using a little notational freedom, we augment $\mathbf{R}$ and A with a special value $\perp$, which will always be discarded by S_0.

In an adaptive integration algorithm, there is also a refinement operation $r : \Gamma \times A \to \Gamma$ that improves the grid based on the auxiliary information $a \in A$. r can be extended naturally to the codomain of S_0

$$\begin{aligned} r : R \times \Gamma \times A \times F \times \mathbf{R} \times \mathbf{R} &\to R \times \Gamma \times A \times F \times \mathbf{R} \times \mathbf{R} \\ (\rho, G, a, f, \mu_1, \mu_2) &\mapsto (\rho, G', a, f, \mu_1, \mu_2) = r(\rho, G, a, f, \mu_1, \mu_2)\,, \end{aligned} \tag{4.54}$$

so that $S = rS_0$ is well defined and we can specify n-step adaptive sampling as

$$S_n = S_0(rS_0)^n \tag{4.55}$$

Since, in a typical application, only the estimate of the integral and the standard deviation are used, a projection can be applied to the result of S_n:

$$\begin{aligned} P : R \times \Gamma \times A \times F \times \mathbf{R} \times \mathbf{R} &\to \mathbf{R} \times \mathbf{R} \\ (\rho, G, a, f, \mu_1, \mu_2) &\mapsto (I, \sigma)\,. \end{aligned} \tag{4.56}$$

Then

$$(I, \sigma) = PS_0(rS_0)^n(\rho, G_0, \perp, f, \perp, \perp) \tag{4.57}$$

and a good refinement prescription r, such as `VEGAS`, will minimize the σ.

For parallelization, it is crucial to find a division of S_n or any part of it into *independent* pieces that can be evaluated in parallel. In order to be effective, r has to be applied to *all* of a and therefore a synchronization of G before and after r is appropriate. Furthermore, r usually uses only a tiny fraction of the overall CPU time and it makes little sense to invest a lot of effort into parallelizing it. On the other hand, S_0 can be parallelized naturally, because all operations are linear, including he computation of a. We only have to make sure that the cost of communicating the results of S_0 and r back and forth during the computation of S_n do not offset any performance gain from parallel processing.

When we construct a decomposition of S_0 and proof that it does not change the results, i.e.

$$S_0 = \iota S_0 \phi \tag{4.58}$$

or

$$
\begin{array}{ccc}
\bigoplus_{i=1}^N G_i & \xrightarrow{\bigoplus_{i=1}^N S_0} & \bigoplus_{i=1}^N G_i \\
\phi \big\uparrow & & \big\downarrow \iota \\
G & \xrightarrow{S_0} & G
\end{array} , \tag{4.59}
$$

where ϕ is a forking operation and ι is a joining operation, we are faced with the technical problem of a parallel random number source ρ.

As made explicit in (4.52), S_0 changes the state of the random number generator ρ. Demanding *identical* results, imposes therefore a strict ordering on the operations and defeats parallelization. In principle, it is possible to devise implementations of S_0 and ρ that circumvent this problem by distributing subsequences of ρ in such a way among processes that the results do not depend on the number of parallel processes.

However, a reordering of the random number sequence will only change the result by an amount on the order of the statistical sampling error, as long as the scale of the allowed reorderings is *bounded* and much smaller than the period of the random number generator[3]. Below, we will therefore use the notation $x \approx y$ for "equal for an appropriate finite reordering of the ρ used in calculating x and y". For our purposes, the relation $x \approx y$ is strong enough and allows simple and efficient implementations.

4.4.3 Multilinear Structure of the Sampling Algorithm

Since S_0 is essentially a summation, it is natural to expect a linear structure

$$
\bigoplus_i S_0(\rho_i, G_i, a_i, f, \mu_{1,i}, \mu_{2,i}) \approx S_0(\rho, G, a, f, \mu_1, \mu_2) \tag{4.60a}
$$

where

$$
\rho = \bigoplus_i \rho_i \tag{4.60b}
$$

$$
G = \bigoplus_i G_i \tag{4.60c}
$$

$$
a = \bigoplus_i a_i \tag{4.60d}
$$

[3] Arbitrary reorderings on the scale of the period of the pseudo random number generator have to be forbidden, of course.

$$\mu_n = \bigoplus_i \mu_{n,i} \tag{4.60e}$$

for appropriate definitions of "$\oplus$". For the moments, we have standard addition

$$\mu_{n,1} \oplus \mu_{n,2} = \mu_{n,1} + \mu_{n,2} \tag{4.61}$$

and since we only demand equality up to reordering, we only need that the ρ_i are statistically independent. This leaves us with G and a and we have to discuss importance sampling and stratified sampling separately.

Importance Sampling

In the cases of naive Monte Carlo without grid optimization and of importance sampling, the natural decomposition of G is to take j copies of the same grid G/j that are identical to G, each with one j-th of the total sampling points. As long as the a are linear themselves, we can add them up just like the moments

$$a_1 \oplus a_2 = a_1 + a_2 \tag{4.62}$$

and we have found a decomposition (4.60). In the case of `VEGAS`, the a_i are sums of function values at the sampling points. Thus they are obviously linear and this approach is applicable to `VEGAS` in the importance sampling mode.

Stratified Sampling

The situation is more complicated in the case of stratified sampling. The first complication is that in pure stratified sampling there are only two sampling points per cell. Splitting the grid in two pieces as above can provide only a very limited amount of parallelization. The second complication is that the a are no longer linear, since they correspond to a sampling of the variance per cell and no longer of the function values themselves.

However, as long as the samplings contribute to disjoint bins only, we can still "add" the variances by combining bins. The solution is therefore to divide the grid into disjoint bins along the divisions of the stratification grid and to assign a set of bins to each processor.

Finer decompositions will incur higher communications costs and other resource utilization. An implementation based on PVM is described in [82], which minimizes the overhead by running identical copies of the grid G on each processor. Since most of the time is usually spent in function evaluations, it makes sense to run a full S_0 on each processor, skipping function evaluations everywhere but in the region assigned to the processor. This is a neat trick, which is unfortunately tied to the computational model of message passing systems such as PVM and MPI. More general paradigms can not be supported since the separation of the state for the processors is not explicit (it is implicit in the separated address space of the PVM or MPI processes).

However, it is possible to implement (4.60) directly in an efficient manner. This is based on the observation that the grid G used by `VEGAS` is factorized into divisions D^j for each dimension

$$G = \bigotimes_{j=1}^{n_{\text{dim}}} D^j \tag{4.63}$$

and that decompositions of the D^j induce decompositions of G

$$\begin{aligned}G_1 \oplus G_2 &= \left(\bigotimes_{j=1}^{i-1} D^j \otimes D_1^i \otimes \bigotimes_{i=j+1}^{n_{\text{dim}}} D^j \right) \oplus \left(\bigotimes_{j=1}^{i-1} D^j \otimes D_2^i \otimes \bigotimes_{i=j+1}^{n_{\text{dim}}} D^j \right) \\ &= \bigotimes_{j=1}^{i-1} D^j \otimes \left(D_1^i \oplus D_2^i\right) \otimes \bigotimes_{j=i+1}^{n_{\text{dim}}} D^j \,. \end{aligned} \tag{4.64}$$

We can translate (4.64) directly to code that performs the decomposition $D^i = D_1^i \oplus D_2^i$ discussed below and simply duplicates the other divisions $D^{j \neq i}$. A decomposition along multiple dimensions is implemented by a recursive application of (4.64).

In `VEGAS`, the auxiliary information a inherits a factorization similar to the factorization (4.63) of the grid

$$a = (d^1, \ldots, d^{n_{\text{dim}}}) \tag{4.65}$$

but not a multilinear structure. Instead, *as long as the decomposition respects the stratification grid*, we find the in place of (4.64)

$$a_1 \oplus a_2 = (d_1^1 + d_2^1, \ldots, d_1^i \oplus d_2^i, \ldots, d_1^{n_{\text{dim}}} + d_2^{n_{\text{dim}}}) \tag{4.66}$$

with "+" denoting the standard addition of the bin contents and "$\oplus$" denoting the aggregation of disjoint bins. If the decomposition of the division would break up cells of the stratification grid, the equation (4.66) would be incorrect, because, as discussed above, the variance is not linear.

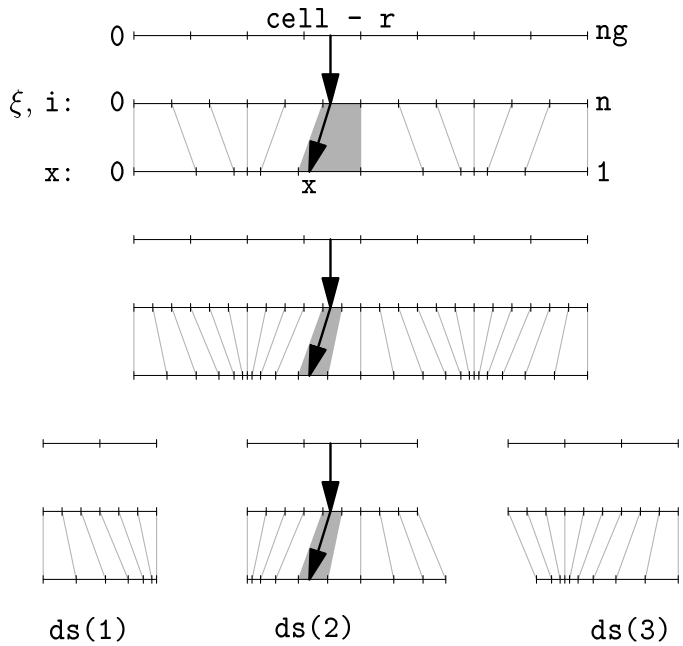


Figure 4.15: *Forking one dimension* **`d`** *of a grid into three parts* **`ds(1)`**, **`ds(2)`**, *and* **`ds(3)`**. *The picture illustrates the most complex case of pseudo stratified sampling (see figure 4.13).*

Now it remains to find a decomposition

$$D^i = D^i_1 \oplus D^i_2 \tag{4.67}$$

for both the pure stratification mode and the pseudo stratification mode of `VEGAS` (see figures 4.11 and 4.12). In the pure stratification mode, the stratification grid is strictly finer than the adaptive grid and we can decompose along either of them immediately. Technically, a decomposition along the coarser of the two is straightforward. Since a typical adaptive grid already has more than 25 bins, a decomposition along the stratification grid offers no advantages for medium scale parallelization ($O(10)$ to $O(100)$ processors) and the decomposition along the adaptive grid has been implemented. This scheme is particularly convenient, because the sampling algorithm S_0 can be applied *unchanged* to the individual grids resulting from the decomposition.

For pseudo stratified sampling (see figure 4.13), the situation is more complicated, because the adaptive and the stratification grid do not share bin boundaries. Since `VEGAS` does *not* use the variance in this mode, it would be theoretically possible to decompose along the adaptive grid and to mimic the

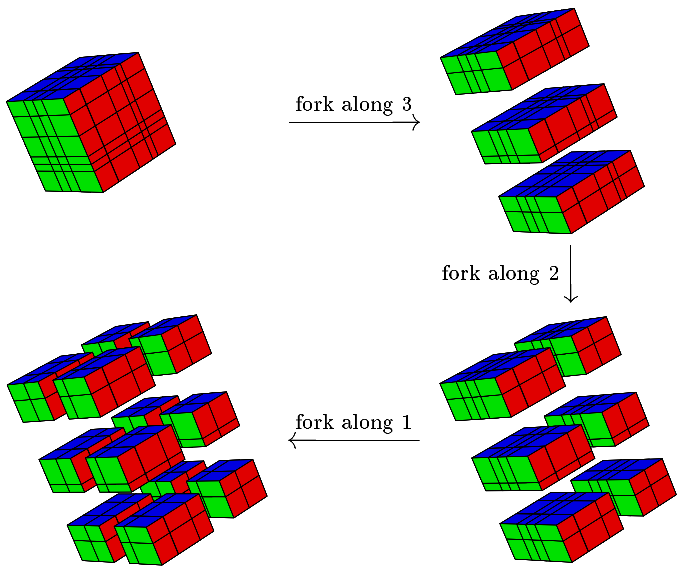


Figure 4.16: *Recursively fork a* $6 \times 5 \times 4$*-grid for distribution among* $12 = 3 \times 2 \times 2$ *processors.*

incomplete bins of the stratification grid in the sampling algorithm. However, this would cause technical complications, destroying the universality of S_0. Instead, the adaptive grid is subdivided in a first step in

$$\mathrm{lcm}\left(\frac{\mathrm{lcm}(n_f, n_g)}{n_f}, n_x\right) \tag{4.68}$$

bins[4]. such that the adaptive grid is strictly finer than the stratification grid. This procedure is shown in figure 4.15. The code for forking and duplicating a single dimension suffices as building block for general recursion [79], as illustrated in figure 4.16.

4.4.4 Random Numbers

The implementation [79] of this procedure takes advantage of the ability of Knuth's new preferred generator [83] to generate *provable* statistically independent subsequences. However, since the state of the random number generator is explicit in all procedure calls, other means of obtaining subsequences can be implemented in a trivial wrapper.

[4]The coarsest grid covering the division of n_g bins into n_f forks has $n_g/\gcd(n_f, n_g) = \mathrm{lcm}(n_f, n_g)/n_f$ bins per fork.

The results of the parallel example will depend on the number of processors, because this effects the subsequences being used. Of course, the variation will be compatible with the statistical error. It must be stressed that the results are deterministic for a given number of processors and a given set of random number generator seeds. Since parallel computing environments allow to fix the number of processors, debugging of exceptional conditions remains therefore possible.

4.5 Dedicated Phase Space Algorithms

No description of Monte Carlo algorithms for phase space integration would be complete without mentioning the beautiful *RAMBO* algorithm [84] for massless n-particle phase space. RAMBO proceeds by generating massless momentum vectors isotropically, without regard to energy or momentum conservation. In a second step, these vectors are boosted with a Lorentz transformation to their common rest frame to enforce momentum conservation and the lengths of massless vectors can be rescaled by a common factor to ensure energy conservation without spoiling momentum conservation. It can be shown [84] that this transformation has a unit Jacobian. Therefore the phase space distribution can be generated without a rejection step. The only obvious additional cost is that a n-particle final state uses $4n$ random numbers for $3n - 4$ degrees of freedom.

Unfortunately, the transformation from the $4n$ random numbers to the $3n-4$ independent variables has a substantial hidden cost: the relation of the integration variables to the kinematical variables in which matrix elements have singularities is lost and adaptive Monte Carlo algorithms are almost useless. For this reason, RAMBO is mainly useful for smooth background distributions. There is a modification (called *MAMBO*) for light massive particles [85], but the Jacobians are now no longer unity and the resulting event weights make the algorithm not very efficient.

—5—
Linear Colliders

Wie Böhmen sie bestand, das einen
schönen Tags zum Meer begnadigt wurde
und jetzt am Wasser liegt.
[Ingeborg Bachmann: *Böhmen liegt am Meer*]

The marquee measurement in the area of gauge boson physics at a linear collider is without a doubt the precision determination of their self-couplings, known as Triple Gauge Couplings (TGC). As has been stressed in chapter 2, the measurements a LEP2 [6] (see figure 1.4), have been a very important confirmation, but can not have been surprising, because a consistent description of the low energy data leaves little room for deviations from the SM predictions at a level that would be detectable at LEP2 [18].

As will be shown in section 5.1, this is qualitatively different at a LC, because higher energy and luminosity allow to probe the couplings at a level where deviations due to different realizations of EWSB are possible.

5.1 Triple Gauge Couplings

The most popular parametrization of the most general Lorentz invariant $W^+W^-Z^0$- and $W^+W^-\gamma$-vertices has been proposed by Hagiwara et al. [86]

$$\begin{aligned}
L_{WWV}/g_{WWV} = {} & \\
& ig_1^V \left(W^\dagger_{\mu\nu}W^\mu V^\nu - W^\dagger_\mu V_\nu W^{\mu\nu}\right) + i\kappa_V W^\dagger_\mu W_\nu V^{\mu\nu} + i\frac{\lambda_V}{M_W^2} W^\dagger_{\lambda\mu} W^\mu{}_\nu V^{\nu\lambda} \\
& - g_4^V W^\dagger_\mu W_\nu \left(\partial^\mu V^\nu + \partial^\nu V^\mu\right) + g_5^V \epsilon^{\mu\nu\lambda\rho}\left(W^\dagger_\mu \partial_\lambda W_\nu - \partial_\lambda W^\dagger_\mu W_\nu\right) V_\rho \\
& + i\tilde\kappa_V W^\dagger_\mu W_\nu \tilde V^{\mu\nu} + i\frac{\tilde\lambda_V}{M_W^2} W^\dagger_{\lambda\mu} W^\mu{}_\nu \tilde V^{\nu\lambda}\,, \qquad (5.1)
\end{aligned}$$

where V is either Z^0 or γ, while $g_{WW\gamma} = -e$ and $g_{WWZ} = -e\cot\theta_w$ are the SM couplings. All field-strengths tensors in (5.1) are abelian $V_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu$ with the tilde denoting the dual tensor. At tree level in the SM, the parameters in (5.1) take the values

$$\Delta\kappa_{\gamma,Z} = \kappa_{\gamma,Z} - 1 = 0 \tag{5.2a}$$
$$\Delta g_1^{\gamma,Z} = g_1^{\gamma,Z} - 1 = 0 \tag{5.2b}$$
$$\lambda_{\gamma,Z} = 0 \tag{5.2c}$$
$$\tilde{\lambda}_{\gamma,Z} = \tilde{\kappa}_{\gamma,Z} = g_4^{\gamma,Z} = g_5^{\gamma,Z} = 0\,. \tag{5.2d}$$

Electromagnetic $\mathrm{U}(1)_Q$ gauge invariance dictates $g_{1,5}^\gamma = 0$ at zero momentum transfer. The parametrization (5.1) is calculationally convenient, because the parameters translate directly to coefficients of the same size in the Feynman rules. Therefore, the matrix elements used in most Monte Carlo event generators are expressed in terms of (5.1).

Another advantage of the parameters in (5.1) is that they translate immediately, without unnaturally small or large coefficients, to model independent physical observables for the electromagnetic properties of charged vector bosons, such as the magnetic dipole

$$\mu_W = \frac{e}{2m_W}\left(g_1^\gamma + \kappa_\gamma + \lambda_\gamma\right), \tag{5.3a}$$

electric quadrupole

$$Q_W^e = -\frac{e}{m_W^2}\left(\kappa_\gamma - \lambda_\gamma\right), \tag{5.3b}$$

electric dipole

$$d_W = \frac{e}{2m_W}\left(\tilde{\kappa}_\gamma + \tilde{\lambda}_\gamma\right), \tag{5.3c}$$

and magnetic quadrupole

$$Q_W^m = -\frac{e}{m_W^2}\left(\tilde{\kappa}_\gamma - \tilde{\lambda}_\gamma\right) \tag{5.3d}$$

moments of $W^\pm$ bosons [87].

On the other hand, (5.1) does not offer any indication to the probable size of the individual contributions. Completely model independent fits of all 14 parameters simultaneously are not possible, even at a high luminosity LC. Furthermore, the parameters in (5.1) are not necessarily constant and could

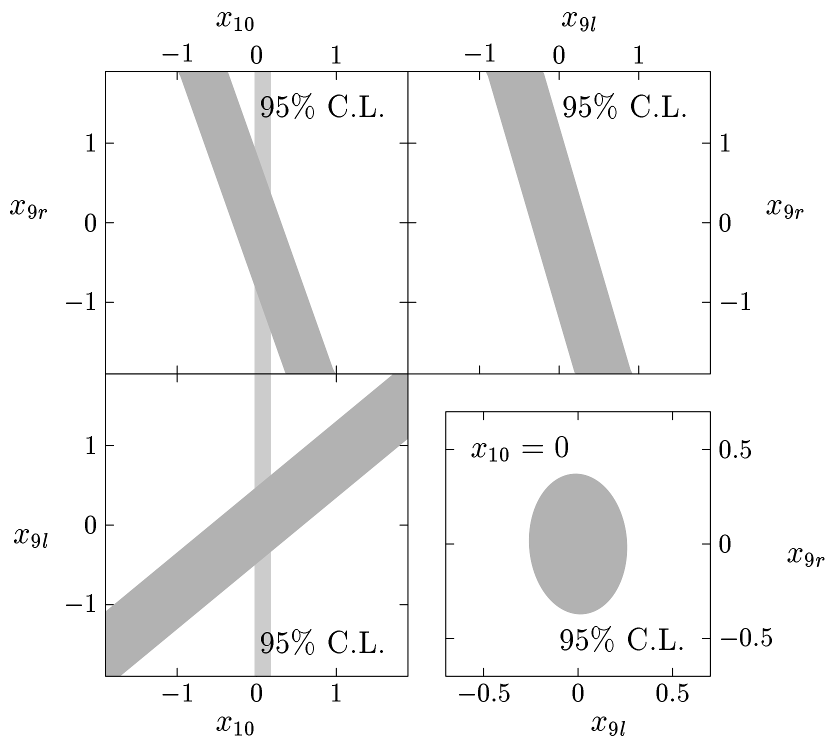


Figure 5.1: *Sensitivity on the parameters of a non-linearly realized EWSB sector with* $\int \mathcal{L} = 200\,\mathrm{fb}^{-1}$ *and* $\sqrt{s} = 800\,\mathrm{GeV}$ *[89]. The vertical bars indicate the constraint on* x_{10} *derived from LEP1 analyses [90]. The blind direction* $(2\sin^2\theta_w, -2\cos^2\theta_w, \cos^2\theta_w - \sin^2\theta_w)$ *does not contribute to the triple gauge couplings.*

be polynomials in the invariant masses or even general form factors. Obviously, g_5^V, $\tilde{\kappa}_V$, and $\tilde{\lambda}_V$ are CP-violating couplings and their study could be postponed.

A more systematic approach starts from a basis [88] for the operators that can appear in an $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y/\mathrm{U}(1)_Q$ effective Lagrangian and classifies them according to their order in the momentum expansion. For non-linear realizations, NDA provides the necessary hierarchy. A systematic study can then start from the lowest dimensional operators and can incorporate symmetries like the custodial $\mathrm{SU}(2)_c$, which is strongly suggested by the low energy electroweak precision data.

The operators of dimension four in the momentum expansion contributing to the TGC are [91]

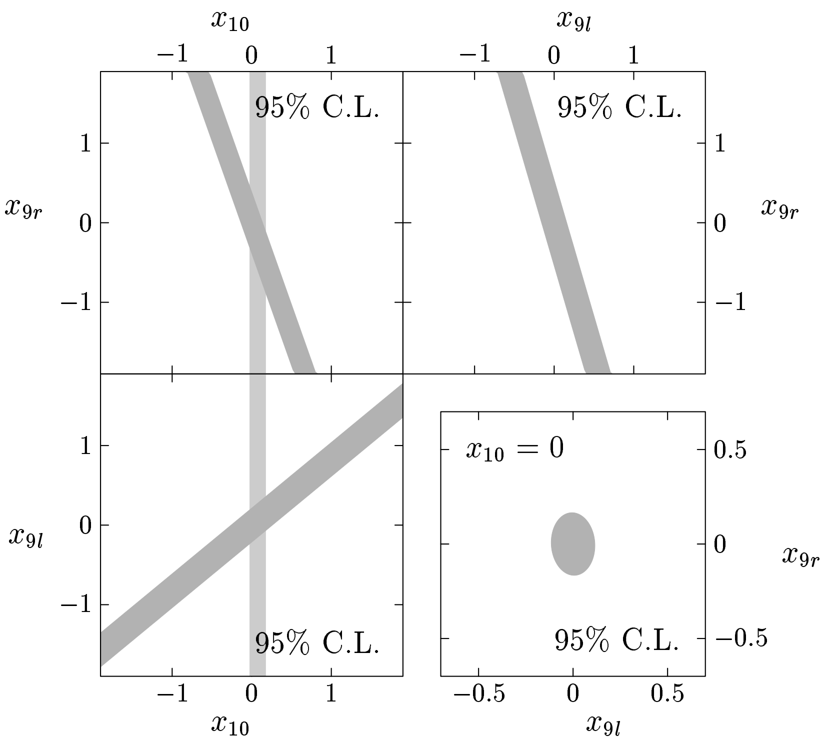


Figure 5.2: *The sensitivity ellipses of figure 5.1 rescaled to the new high luminosity TESLA with* $\int \mathcal{L} = 1\,\mathrm{ab}^{-1}$.

$$
\begin{aligned}
L = &-i\frac{x_{9l}}{16\pi^2}\,\mathrm{tr}\Big(gW^{\mu\nu}D_\mu U^\dagger D_\nu U\Big) \\
&- i\frac{x_{9r}}{16\pi^2}\,\mathrm{tr}\Big(g'B^{\mu\nu}D_\mu U^\dagger D_\nu U\Big) + \frac{x_{10}}{16\pi^2}\,\mathrm{tr}\Big(U^\dagger g' B_{\mu\nu} U g W^{\mu\nu}\Big)\,. \qquad (5.4)
\end{aligned}
$$

The contributions of the interaction (5.4) to the TGC can be expressed in terms of the general parametrization (5.1) by comparing coefficients[1] of the trilinear terms

$$g_1^Z = 1 - \frac{e^2}{2\sin^2\theta_w\cos^2\theta_w}\frac{x_{9l}}{16\pi^2} + \frac{e^2}{\cos^2\theta_w(\cos^2\theta_w - \sin^2\theta_w)}\frac{x_{10}}{16\pi^2} \qquad (5.5a)$$

$$\kappa_Z = 1 - \frac{e^2}{2\sin^2\theta_w}\frac{x_{9l}}{16\pi^2} + \frac{e^2}{2\cos^2\theta_w}\frac{x_{9r}}{16\pi^2} + \frac{2e^2}{\cos^2\theta_w\sin^2\theta_w}\frac{x_{10}}{16\pi^2} \qquad (5.5b)$$

$$\kappa_\gamma = -\frac{e^2}{2\sin^2\theta_w}\frac{x_{9l}}{16\pi^2} - \frac{e^2}{2\sin^2\theta_w}\frac{x_{9r}}{16\pi^2} - \frac{e^2}{\sin^2\theta_w}\frac{x_{10}}{16\pi^2}\,. \qquad (5.5c)$$

[1] The erroneous non-singular transformation formula in [91] is corrected for the special case $x_{9l} = x_{9r}$ in [92] and for the general case in [93].

Unfortunately, the transformation (5.5) is singular and the blind direction

$$\begin{pmatrix} x_{9l} \\ x_{9r} \\ x_{10} \end{pmatrix}_{\text{blind}} \propto \begin{pmatrix} 2\sin^2\theta_w \\ -2\cos^2\theta_w \\ \cos^2\theta_w - \sin^2\theta_w \end{pmatrix} \tag{5.6}$$

in the parameter space can not be observed in measurements of the TGC alone. However, LEP1 analyses (e. g. [90]) provide independent constraints on x_{10}, which can lift the degeneracy.

The coefficients in (5.4) have been normalized such that their natural size according to NDA (see sections 2.7 and A.3.5) is of order one. Consequently, the all important question from a physics perspective is: *Are experiments at a LC sensitive to values for x_{9l}, x_{9r}, x_{10} of order one?*

Detector effects appear to be small and preliminary studies can therefore work on the generator level with simple geometrical acceptance cuts [94]. Much more important than detector effects are the distortions of the distributions from hard, almost collinear initial state photon radiation, when the photon is lost in the beam-pipe. Since these effects cause a migration of events in the phase space, simple acceptance corrections do not suffice and more complicated unfolding procedures are required.

In one study [89] (see also [3]) the unfolding has been performed by generating a large sample of Monte Carlo events (ten times the size of the expected data sample at TESLA) with `WOPPER` [59] and summing over all radiative and non-radiative events with the same experimental signature. The events were subsequently reweighted with anomalous TGC using fast matrix elements of `WPHACT` [95] to obtain the likelihood function for the TGC. Using a quadratic approximation of the likelihood function for the "old" high luminosity option at TESLA with $\int\mathcal{L} = 200\,\text{fb}^{-1}$ and $\sqrt{s} = 800\,\text{GeV}$, the sensitivities in figure 5.1 have been derived. A more careful analysis will have to take non-linearities in the likelihood function into account[2], but they can not change the conclusion that *the LC is definitely sensitive to values for x_{9l}, x_{9r}, x_{10} of order one,* if constraints on x_{10} are taken into account.

Increasing the integrated luminosity according to the "new" TESLA high luminosity option by a factor of five to $\int\mathcal{L} = 1\,\text{ab}^{-1}$, as shown in figure 5.2, shows that such a machine would allow precision tests of the TGC, at the level of 10% of reasonable deviations from the SM, caused by different realizations of the EWSB sector.

[2]The non-linearities had been neglected because the unfolding would have exceeded the limited computing resources in a non-linear fit.

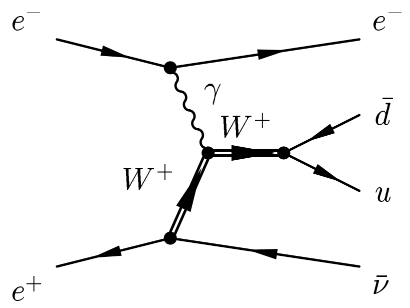


Figure 5.3: *"Single-W" contribution to $e^+e^- \to e^-\bar{\nu}u\bar{d}$.*

5.2 Single $W^\pm$ Production

At energies that are high compared to the masses of all participating particles, all cross sections with a finite limit for massless particles have to fall off as dictated by dimensional analysis

$$\sigma(s) \propto \frac{1}{s}\,. \tag{5.7a}$$

Cross sections that have contributions from the exchange of light particles in the t-channel do not necessarily have a finite limit and can diverge in the forward direction if the mass of the exchanged particle is send to zero. Such cross sections can scale like

$$\sigma(s) \propto \frac{1}{\mu^2}\ln\left(\frac{s}{\mu^2}\right)\,, \tag{5.7b}$$

where μ^2 denotes a typical kinematical scale set by a minimum momentum transfer.

For the t-channel neutrino exchange in $W^\pm$ pair production, the forward singularity is cut off kinematically by the $W^\pm$-mass and the cross section is still dominated by (5.7a) in the TeV-region. The situation is different for "single-W" production as shown in figure 5.3, because the electron or positron that goes undetected in the forward direction does not provide a kinematical cut-off. The observable cross section is still small at LEP2, but it will be larger at a LC than that of $W^\pm$ pair production.

A unique physics opportunity of single-W production in the context of TGC is that, unlike $W^\pm$ pair production, the diagram in figure 5.3 is sensitive to the $W^+W^-\gamma$-couplings alone, because the Z^0-exchange is not singular. Therefore it can be used to disentangle $W^+W^-Z^0$- and $W^+W^-\gamma$-couplings without studying $\gamma\gamma$-collisions in a dedicated collider.

For LEP2, $W^\pm$ pair production has been studied extensively [96] and many reliable Monte Carlo event generators have been prepared [97]. A more recent dedicated comparison of theoretical predictions for single-W production has revealed that the situation is less favorable for single-W production. At LC energies, unacceptable variations of the predictions for the total cross section on the order of up to five percent have been reported [98, 99]. It appears that these discrepancies can be traced back to incomplete implementations of effects from the finite electron mass.

The lesson to be learned is that a simple translation of the electron mass into a lower cut-off on the electron scattering angle will not suffice at a LC and more complete calculations in the forward region are required, if single-W production is to be used for precision physics, like the disentangling of $W^+W^-Z^0$- and $W^+W^-\gamma$-couplings.

The classification of the gauge invariant classes of Feynman diagrams for single-W production (see figure 3.7) allows to improve the handling of radiative corrections. The single-W and $W^\pm$ pair production contributions to the same four fermion final state can be folded with a separate radiator function, without spoiling gauge invariance. Since the two classes have very different characteristic scales and the interferences are small, this improves the predictions considerably.

5.3 Beamstrahlung

Another issue that differentiates a LC from LEP2 is the non-trivial beam energy spectrum caused by *beamstrahlung.* Unlike initial state radiation, beamstrahlung is not characterized by the scale of the hard interaction, but by the collective interaction of the crossing bunches, which must be much denser than the bunches in storage rings to obtain sufficient luminosity. Therefore, beamstrahlung is a more complex physical phenomenon, but it is independent of the hard interaction under study and can be parametrized in a single universal beam spectrum for each collider design at each beam energy.

The comparison of analytical approximations [100] with the results of microscopic simulations [101, 102] reveals that the former are not adequate. A successful pragmatical solution is to parametrize the results of the microscopic simulations with a simple factorized *ansatz* [103]

$$D_{p_1p_2}(x_1, x_2, s) = d_{p_1}(x_1)d_{p_2}(x_2) \tag{5.8}$$

where $x_{e^\pm,\gamma} = E_{e^\pm,\gamma}/E_{\text{beam}}$ and

$$d_{e^\pm}(x) = a_0\delta(1-x) + a_1 x^{a_2}(1-x)^{a_3} \tag{5.9a}$$

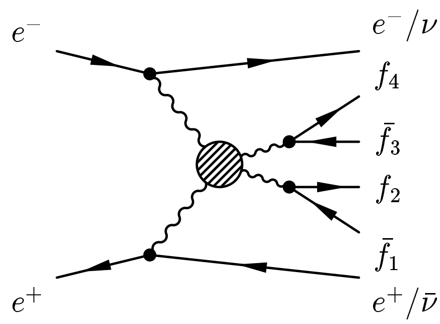


Figure 5.4: *Vector boson scattering subprocess in six-fermion production.*

$$d_\gamma(x) = a_4 x^{a_5} (1-x)^{a_6} . \tag{5.9b}$$

For $a_{0,1,4} \geq 0$ and $a_{2,3,5,6} \geq -1$, the ansatz (5.9) corresponds to a positive and integrable distribution can reproduce the soft singularities from multiple emission at $x_{e^\pm} \to 1$ and $x_\gamma \to 0$ for $a_{3,5} < 0$.

The ansatz (5.9) is able to fit the results of microscopic simulations well [103]. Fits for the latest collider parameter sets are available as distributions and as efficient random number generator in the library Kίρκη [104].

5.4 Quadruple Gauge Couplings and $W^\pm$ Scattering

While LEP2 is, for gauge boson physics, essentially a four fermion production machine, a LC can also produce final states with six or more fermions, which have contributions from quadruple gauge boson couplings. In addition to the coupling of an s-channel Z^0 or γ to three vector bosons, there are the kinematical configurations corresponding to the fusion of almost on-shell vector bosons, as shown in figure 5.4. This part of the phase space is particularly attractive for studying the interactions of strongly interacting vector bosons.

Realistic estimates for the physics reach in this channl have already been obtained using an improved equivalent particle approximation for the “production” of the vector bosons in the intermediate state [105] (see also [3]). Interesting physics can be probed, but a high luminosity option is required.

The construction of a complete unweighted event generator for $e^+e^- \to 6f$, including general gauge boson couplings, is still a challenge. Complete calculations for selected channels exist [106, 107], which are expected to be reliable in the regions of the phase space corresponding to top pair production and Higgs production. More work is required for a proper treatment of the part of the phase space that corresponds to the scattering of almost on-shell

$W^{\pm}$'s as in figure 5.4. Currently, work is underway that uses the Monte Carlo methods described in section 4.3.4 to bridge this gap.

—6—
Conclusions

While it is never safe to say that the future of Physical Science has no marvels even more astonishing than those of the past, it seems probable that most of the grand underlying principles have been firmly established and that further advances are to be sought chiefly in the rigorous applications of these principles to all the phenomena which come under our notice. It is here that the science of measurement shows its importance—where quantitative results are more to be desired than qualitative work. An eminent physicist has remarked that the future truths of Physical Science are to be looked for in the sixth place of decimals.
[Albert Michelson]

There is ample model independent evidence from low energy data that weak interactions exchange vector and axial-vector quantum numbers, coupling to charged and neutral currents.

The attempt to incorporate these interactions in a renormalizable QFT in which perturbation theory is well defined without any short distance cut-off has singled out spontaneously broken gauge theories in which the weak interactions are described by the exchange of massive gauge bosons. The same conclusion is reached by attempting to enforce high energy tree-level unitarity of scattering amplitudes.

As discussed in chapter 2 (with a little help from appendix A), gauge theories retain their special status, even if the electroweak SM is interpreted as an EFT, which may have to be replaced by a more comprehensive theory at a energy scale beyond the physics reach of current colliders. Instead of guaranteeing the consistency of formal[1] perturbation theory to arbitrarily high energies, the improved high energy behavior is used for organizing the expansion of the effective interaction in a more systematic way than would be possible with non-gauge vector bosons. Selecting the gauge theory among all equivalent EFTs that describe the same physics (on-shell matrix elements) does not sacrifice any generality.

Chapter 3 described how to systematically organize tree-level calculations for many particle final states in such a way that non-abelian gauge invariance can be maintained by partial sums. This is important in state-of-the-art calculations, that would exhaust available human and computing resources if they had to be done in a single step. In chapter 4, methods for improving numerical convergence and parallelization of computations have been presented, that can be used for obtaining high precision theoretical predictions for event rates in channels with many particles in the final state, as discussed briefly in chapter 5.

Currently, a *pessimistic* scenario can not be excluded, in which a fundamental Higgs particle will be found, together with a few supersymmetry partners of SM particles, compatible with the MSSM. While these discoveries would be a tremendous achievement of experimental physics and the culmination of decades of theoretical work, it would probably put the answers to the open questions in the physics of flavor out of our reach. However, we must not forget which eminent *"future truths of Physical Science"* were found when Michelson had a look at *"the sixth place of decimals."* Rumors about the completion of physics are almost always greatly exaggerated ...

[1] I. e. without regard to the convergence of the perturbative series.

Appendix

—A— The Renormalization Group

Wisse, daß dies Universum das ist, was es zu sein vorgibt: ein Unendliches. Versuche nie im Vertrauen auf deine logische Verdauungskraft, es zu verschlingen; sei vielmehr dankbar, wenn du durch geschicktes Einrammen dieses oder jenes festen Pfeilers in das Chaos verhinderst, daß es dich verschlinge.
[Thomas Carlyle: *The French Revolution*[1]]

Until a few year back, textbooks on elementary particle physics and QFT have hardly mentioned the method of EFT and have usually given treatments of the RG that are have been more formalistic than lucid. Consequently, these subjects have had more of an oral tradition than canonical scriptures. Fortunately this has changed now with the advent of two excellent textbooks offering both a pedagogical [21] and a systematical [19, 20] exposition of QFT.

These lectures have been given to an audience of graduate students working in experiments and on theoretical projects. Therefore they had to live up to the combined challenge of presenting theorists new angles on a subject they had already studied, while at the same time offering concrete applications to experimentalists, without burying them in formalism.

The intuitive reasoning in these lectures avoids the horrendous combinatorial complications of perturbative renormalization theory. Nevertheless, mathematical proofs of the intuitive observations still need the technical results of traditional perturbative renormalization theory.

These lectures deliberately use a strictly perturbative language in order to appeal to the audience's intuition that is trained on Feynman diagrams. Many results remain valid outside of PT, but no attempt is made to mention this systematically.

[1] Translation and quote from Thomas Mann: *Betrachtungen eines Unpolitischen.*

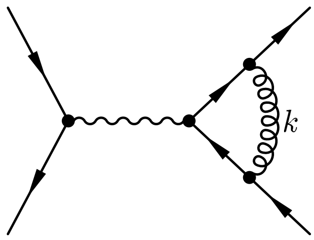


Figure A.1: *One-loop contribution to $e^+e^- \to q\bar{q}$.*

A.1 Introduction

Ein Teil von jener Kraft,
Die stets das Böse will und stets das Gute schafft.
[J. W. v. GOETHE: *Faust I*]

What is the size of leptons and quarks? According to the most recent report of the Particle Data Group (PDG) [8], the compositeness scale is larger than 1.5 TeV for all of them, but even the best limits fall short of 5 TeV. In the SM, we assume all fermions to be strictly pointlike. As you all know, the SM is very successful. Yet, we can't pin down the size of its constituents below $0.2\,\mathrm{TeV}^{-1}$. *How is this possible?*

The easy way out is to argue that our accelerators are too small to probe any deeper. This is *too* easy, however, because we claim to control the SM at the quantum (i. e. one-loop) level. The description of a typical LEP1 experiment $e^+e^- \to q\bar{q}$ involves loop diagrams like the one in figure A.1. And in this diagram, the loop momentum k is *not* bounded anywhere and there is a part of the integration region where it exceeds the energy scale $\Lambda \approx 1/\Delta x$.

At this point we can rephrase the question: *Why is the compositeness scale irrelevant to physics at LEP1 energies?* The answer will be that the compositeness interactions

$$\Delta L = \frac{1}{2\Lambda^2}\,\bar{\psi}\Gamma_\mu\psi\,\bar{\psi}\Gamma^\mu\psi \tag{A.1}$$

are indeed *irrelevant* and the term *"irrelevant"* will be given a precise technical definition below. This might be a bit surprising, because most of us have learned in graduate school that interactions like (A.1) are *verboten* in loop calculations, because they make the theory non-renormalizable.

A closely related question is: *why don't we have to understand M-theory to do phenomenological particle physics?* According to the theoretical party

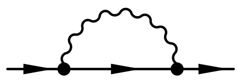


Figure A.2: *Electron self energy*

line, we expect a lot of new particles (grand unification gauge and Higgs bosons, higher string and membrane modes, etc.) somewhere above 10^{16} GeV. Why do they have so little impact on our bread and butter phenomenology that involves loop diagrams?

A.1.1 History

QFT was introduced shortly after Quantum Mechanics (QM) in the 1920s. But while QM was put on a solid mathematical basis in the 1930s, it took much longer to make sense out of QFT.

Soon (in the early 1930s), it was discovered that the second order of PT gave non-sensical results in QED. For example, the electron self energy (see figure A.2, but the calculators in the 1930s had to work without the benefits of Feynman diagrams, of course) came out infinite. Even though the lowest order predictions were very successful, the inability to calculate corrections was unsettling, to put it mildly. By the late 1940s, a workable scheme was established for absorbing the infinities in renormalized couplings constants and masses. Nevertheless, this process of *renormalization* was considered a nuisance and it was expected that a complete theory would not need it.

Even though the renormalization procedure apparently worked in practical applications, it took another twenty years of hard theoretical work until a mathematical proof that the procedure works in all orders of PT was established in the 1960s.

Today, the formalism of perturbative QFT is well established and the electroweak precision measurements at LEP1 leave no doubt that its results are phenomenologically sound. Nevertheless, it should be remembered that no mathematical proof for the existence of an interacting QFT in four space-time dimensions has been given that works outside of the realms of PT.

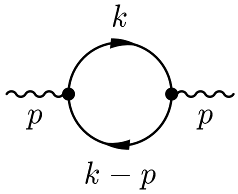


Figure A.3: *Vacuum polarization*

A.1.2 Renormalizability

Let us look at a simple example to get a feel for the physics of renormalization. For the vacuum polarization diagram of QED (see figure A.3)

$$i\Pi_{\mu\nu}(p) = -i(p^2 g_{\mu\nu} - p_\mu p_\nu)\Pi(p^2) \tag{A.2}$$

$$\Pi(p^2) = \frac{\alpha}{3\pi}\int_{-p^2}^{-\infty} \frac{\mathrm{d}(-k^2)}{(-k^2)} + \text{finite} \tag{A.3}$$

we find a logarithmically divergent integral over k. Obviously, the divergence comes from the region $-k^2 \to \infty$, where the $ee\gamma$-vertex has not been tested experimentally. A brute force remedy is to cut-off the integral at an arbitrary, but very large, scale Λ

$$\tilde{\Pi}(p^2) = \frac{\alpha}{3\pi}\int_{-p^2}^{\Lambda^2} \frac{\mathrm{d}(-k^2)}{(-k^2)} + \text{finite} = \frac{\alpha}{3\pi}\ln\frac{\Lambda^2}{-p^2} + \text{finite} \tag{A.4}$$

This leaves us with a logarithmic dependence on the unphysical parameter Λ and is not acceptable in this form. We can however impose a renormalization condition

$$\tilde{\Pi}(p^2)\Big|_{-p^2=\mu^2} = 0 \tag{A.5}$$

that trades the dependence on the *unphysical* cut-off Λ for the dependence on the *physical* renormalization point μ at which we define the charge. Even though the result is now finite

$$\Pi^R(p^2) = -\frac{\alpha}{3\pi}\ln\frac{-p^2}{\mu^2} \tag{A.6}$$

we are faced with a potentially large logarithm $\ln(p^2/\mu^2)$, that reaches 25 for LEP2 energies ($\sqrt{s} \approx 190\,\mathrm{GeV}$) with a charge renormalized at low energies ($\mu \approx m_e$). The small coupling constant of QED helps to keep this

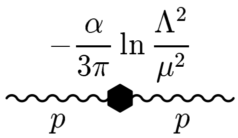


Figure A.4: *Vacuum polarization counterterm*

logarithm relatively harmless, but since LEP1 has introduced the era of electroweak precision physics, higher orders in PT are required for reliable quantitative predictions. We will see below, that there are cases, where the effective coupling is much larger and a resummation of the perturbation series is required.

It is far from obvious that the subtraction procedure, that I have sketched so crudely, works to all orders in PT. It is not even obvious that all required subtractions correspond to a renormalization of coupling constants and masses. While I will not go into detail, it will be helpful later to establish some nomenclature at this point.

The subtraction leading from (A.4) to (A.6) is obviously equivalent to adding the new *counterterm* vertex depicted in figure A.4 to the theory. Thinking in terms of counterterms instead of subtractions has the advantage that the generalization to higher orders is more intuitive.

In order to get a feeling for the required counterterms, let us consider the free field actions for scalars ϕ, spin-1/2 fermions ψ, and vectors A_μ

$$L_0^\phi = \int \mathrm{d}^4x \, \left(\frac{1}{2} \frac{\partial\phi(x)}{\partial x_\mu} \frac{\partial\phi(x)}{\partial x^\mu} - \frac{m_\phi^2}{2}\phi^2(x) \right) \tag{A.7a}$$

$$L_0^\psi = \int \mathrm{d}^4x \, \left(\bar\psi(x) i\gamma_\mu \frac{\partial}{\partial x_\mu}\psi(x) - m_\psi \bar\psi(x)\psi(x) \right) \tag{A.7b}$$

$$L_0^A = \int \mathrm{d}^4x \, \frac{-1}{4} F_{\mu\nu}(x) F^{\mu\nu}(x) \,, \quad F_{\mu\nu}(x) = \frac{\partial A_\nu(x)}{\partial x^\mu} - \frac{\partial A_\mu(x)}{\partial x^\nu} \,. \tag{A.7c}$$

We are using units with $\hbar = c = 1$ and therefore the actions must be dimensionless. If we assign dimension $+1$ to mass

$$\dim(m) = 1 \,, \tag{A.8}$$

then length has dimension -1 in these units

$$\dim\left(\mathrm{d}^4x\right) = -4 \tag{A.9a}$$

$$\dim\left(\frac{\partial}{\partial x_\mu}\right) = 1 \,. \tag{A.9b}$$

Now we can read off the dimensions of the fields from (A.7) immediately

$$\dim\left(\phi(x)\right) = 1 \tag{A.10a}$$

$$\dim\left(\psi(x)\right) = \frac{3}{2} \tag{A.10b}$$

$$\dim\left(A_\mu(x)\right) = 1\,. \tag{A.10c}$$

These dimensions are reflected in the high energy behavior of the propagators, which is $1/p^{2d-4}$

$$\int \mathrm{d}^4x\, e^{ipx} \left\langle 0\middle|\mathrm{T}\phi(x)\phi(0)\middle|0\right\rangle = \frac{i}{p^2 - m_\phi^2 + i\epsilon} \tag{A.11a}$$

$$\int \mathrm{d}^4x\, e^{ipx} \left\langle 0\middle|\mathrm{T}\psi(x)\bar\psi(0)\middle|0\right\rangle = i\frac{\not{p} + m_\psi}{p^2 - m_\psi^2 + i\epsilon} \tag{A.11b}$$

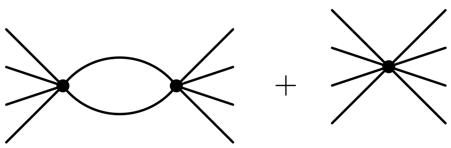
$$\int \mathrm{d}^4x\, e^{ipx} \left\langle 0\middle|\mathrm{T}A_\mu(x)A_\mu(0)\middle|0\right\rangle = \frac{-ig_{\mu\nu}}{p^2 + i\epsilon}\,. \tag{A.11c}$$

The power counting for loop integrals in Feynman diagrams can be related directly to the dimensions of the interaction operators. Let me illustrate this by a some examples[2]. I will give a more comprehensive account from the point of view of the RG in later sections.

Consider a one-loop integral with two ϕ^6 insertions. Since the loop integral

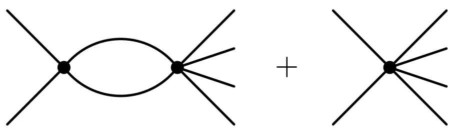
$$\int \frac{\mathrm{d}^4k}{(2\pi)^4}\,\frac{1}{k^2(p-k)^2} \tag{A.12}$$

is logarithmically divergent, we need a ϕ^8 counterterm



$$= \text{finite}\,. \tag{A.13}$$

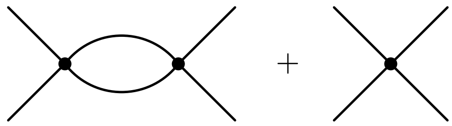
Similarly, a ϕ^4 and a ϕ^6 operator need a ϕ^6 counterterm



$$= \text{finite} \tag{A.14}$$

[2] A systematic exposition too all orders in PT is well beyond the scope of these lectures, but can be found in any QFT text.

and two ϕ^4 operators need a ϕ^4 counterterm



$$= \text{finite}\,. \tag{A.15}$$

This suggests a pattern, which can indeed be verified by a more detailed analysis. If we have multiple insertions of operators with dimension greater than four, then counterterms of higher and higher dimension are needed. On the other hand, operators of dimension less than or equal to four require counterterms of dimension less than or equal to four.

This pattern can already be seen from dimensional arguments. If an operator has dimension greater than four, it must come with a coupling constant of negative dimension, e. g.

$$\frac{1}{\Lambda^2}\frac{1}{6!}\phi^6(x)\,,\quad \dim(\Lambda) = 1\,. \tag{A.16}$$

If the loop integral does not depend on the scale in this coupling constant, then the product of two such operators will have a coupling constant of an even more negative dimension. The product will therefore be of even higher dimension, e. g.

$$\frac{1}{\Lambda^2}\frac{1}{6!}\phi^6(x)\frac{1}{\Lambda^2}\frac{1}{6!}\phi^6(y) \to \frac{1}{16\pi^2}\frac{1}{\Lambda^4}\frac{1}{8!}\phi^8(x)\,. \tag{A.17}$$

Since all building blocks have a positive dimension (see (A.9b) and (A.10)), there are only a finite number of operators of a given dimension. Therefore, interactions of dimension less then four can be renormalized by a *finite* number of counterterms[3]. As long as a *finite* number of counterterms suffices, the theory retains its predictive power. As a result, interactions of dimension four are called *renormalizable* and interactions of dimension less than four are called *super-renormalizable*. There are additional complications with the preservation of symmetries, which do not affect the general argument, however.

It was then very reasonable to expect that nature is described by a renormalizable QFT, such as QED, QCD or the SM. These theories have been very successful and as a result, all other—*non-renormalizable*—theories fell from grace and became second class citizens.

However, two very important questioned were apparently never posed, let alone answered:

[3] While this is rather intuitive from the one-loop examples, the mathematical proof is extremely complicated in the general case. It took forty years of efforts [52] of some of the brightest theoretical physicists of their time to disentangle the multi-loop calculations when the loops are not simply nested, but are *overlapping*.

1. *Why is nature described by a renormalizable QFT?* I will show later that the RG will provide a satisfactory answer to this question.

2. *Why should our low energy QFT be correct to arbitrarily high energies?* The contribution of string theory to particle physics that this question is no longer taboo.

A.1.3 The Renormalization Group

The great Cinderella story of theoretical particle physics since 1970 has been the rise of renormalization from a nuisance *(sweeping things under the rug)* to a powerful tool. The turning points of this epic are

1971: Wilson's RG [22] for critical phenomena (i. e. phase transitions) and strong interactions proved that renormalization could help physics insight, instead of hindering it.

1972: The proof of renormalizability of spontaneously broken gauge theories [50, 108] established a viable candidate for a QFT of weak interactions (now known as SM).

1973: Asymptotic freedom [109] paved the way for QCD, the QFT of strong interactions.

1979: Weinberg's EFT paper [110] summarized and formalized the folklore of systematic expansions in the energy.

1990s: Even the non-renormalizable theories have been reinstated as first class citizens under the name of EFT.

By the time of Weinberg's paper, the physics and the formalism was well understood by the experts, but it remained for some time part of the oral tradition that was covered insufficiently by textbooks.

A.2 The Renormalization Group

caminante, no hay camino
se hace camino al andar.
[ANTONIO MACHADO: *Proverbios y cantares, VI*]

We evaluate the predictions of QFT by doing integrals: loop integrals from Feynman diagrams in PT and Feynman path integrals beyond PT. The RG has a very intuitive interpretation [111] (see also [112]) in the evaluation of Feynman's path integral

$$\int \mathcal{D}\phi \, e^{iS(\phi)} \tag{A.18}$$

and I encourage everyone to consult the reference. But I can not assume a previous exposure to (A.18) and a detailed discussion is beyond the scope of these lectures. Therefore I will concentrate on Feynman diagrams and PT.

A.2.1 Cracking Integrals

There are (at least) three common ways to do integrals in physics: either from the anti-derivative or from the theorem of residues or from a Riemann- (or Cauchy- or Lebesgue- or ...) construction. The first method is familiar to all of you from high school days, but it is frowned upon in certain circles because the second method enables physicists to do "harder" integrals and the third method allows mathematicians to integrate more (i. e. "wilder") functions. As one might guess from this introduction, I will now sing the praise of the first method.

The problem of calculating the integral

$$I(f;a,b) = \int_a^b \mathrm{d}x \, f(x) \tag{A.19}$$

is equivalent to solving the first order Ordinary Differential Equation (ODE)

$$\frac{\mathrm{d}}{\mathrm{d}x} F(x) = f(x), \quad F(b) = C\,, \tag{A.20}$$

since the integral is given by $I(f;a,b) = F(a) - C$. This relation is familiar, because it is used in simple cases in the opposite direction to obtain solutions to (A.20) from (A.19). In fact, we will use it frequently below.

Solving the ODE (A.20) might appear more difficult than doing the integral (A.19) and this is true in the general case. However, I will try to convince you that, for many applications in particle physics, the ODE can give us very reliable approximate results, where the value of the corresponding integral is hard (or even impossible) to estimate. In addition, the ODE can give us more physical insight than brute force estimates of integrals. I will also argue that there are cases, where the integral (A.19) does not exist, while the ODE (A.20) can still give us the physically correct answer.

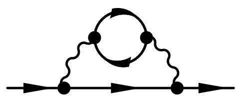


Figure A.5: *Nested divergencies*

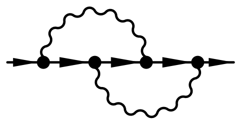


Figure A.6: *Overlapping divergencies*

A.2.2 Wilson's Insight

We have seen heuristically in section A.1 that the divergencies and large logarithms in PT originate from logarithmic integrals in loop momenta. For *nested divergencies* (see figure A.5), it is intuitively clear that the proper procedure is to proceed from the inside out and to subtract the subdivergencies first. Then only the overall divergency from the outermost loop remains.

More intricate is the treatment of *overlapping divergencies* (see figure A.6), where each subdivergency has to be subtracted without double counting. The proof that the subtraction algorithm is mathematically sound to all orders in PT is an extremely complicated combinatorial exercise [52].

Nevertheless, it can be given a lucid physical interpretation, if we change our point of view. Why should we insist of doing the loop integrals first and subtract divergencies later? As we have seen in (A.4) and (A.6), the leading logarithmic dependence on the renormalization scale can be recovered from the dependence on the cut-off by dimensional analysis. Wouldn't it be easier to calculate the dependence on the cut-off directly?

As we shall see now, this heuristic argument can be made precise and results in a powerful calculational tool, the RG. The essence of the method lies in deriving a differential equation, the Renormalization Group Equation (RGE), that can be integrated to give the desired result. It will turn out that the calculation never involves any infinite numbers and in most cases, the coefficients of the PT will never be large.

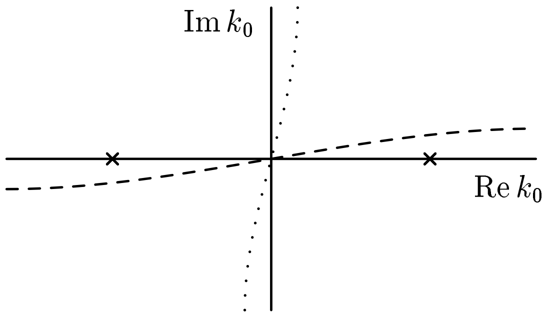


Figure A.7: *Wick rotation*

A.2.3 The Sliding Cut-Off

For simplicity, I will now concentrate on massless ϕ^4-theory. This is the theory of a spinless, uncharged, massless particle, that interacts through a local ϕ^4 vertex. The arguments will go through for more realistic QFTs, but the technical complications arising from gauge invariance would obscure the point that I'm trying to make. I will deal later with the effects from massive particles.

Let us assume for the moment that the theory has been defined with a cut-off in momentum space, i. e. a propagator

$$iD_\Lambda(k^2) = \frac{i}{k^2 + i\epsilon}\Theta(|k| \leq \Lambda)\,. \tag{A.21}$$

The cut-off $|k| \leq \Lambda$ needs some explanation, because it appears not to be Lorentz invariant. In fact, $\Theta(|k| \leq \Lambda)$ is just a symbolic notation. A simple minded cut-off like $|k^2| \leq \Lambda^2$ is invariant, but it is not restrictive enough, because k can grow along light-like directions ($k^2 = 0$) without being cut off.

A more appropriate definition can be found by using the "Wick rotation". As can be seen from figure A.7, in each loop integration, the integration contour in the complex k_0 plane can be deformed from the dashed curve to the dotted curve without crossing singularities. Then we can perform the substitution $(k^0, \vec{k}) = (ik_E^0, \vec{k}_E)$ and the Minkowski "length" becomes a Euclidean length $k^2 = (k^0)^2 - \vec{k}^2 = -(k_E^0)^2 - \vec{k}^2 = -k_E^2$. The Lorentz invariant cut-off $k_E^2 \leq \Lambda^2$ is now effective, because there are no light-like directions in Euclidean space. We shall interpret $\Theta(|k| \leq \Lambda)$ in this fashion.

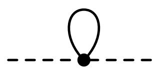
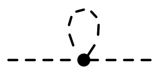
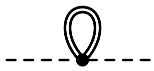
We can now compare the theories defined with two different cut-offs Λ and $\Lambda' < \Lambda$. It turns out that we can change the theory in such a way that the physics does *not* change when we lower the cut-off, as long as the external momenta remain below the new cut-off. To see how this works, let us introduce a graphical notation

$$iD_\Lambda(k^2) = \text{———} \tag{A.22a}$$

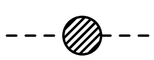
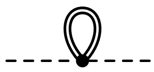
$$iD_{\Lambda'}(k^2) = \text{- - - - - - -} \tag{A.22b}$$

$$iD_\Lambda(k^2) - iD_{\Lambda'}(k^2) = \text{═══} \tag{A.22c}$$

for propagators in the full theory, the low energy theory and their difference. For the one-loop correction to the propagator we find below the new cut-off

 $=$  $+$  . (A.23)

Therefore the physics will not change, if we introduce a new vertex

 $=$  (A.24)

into the low energy theory. When we calculate (A.24) explicitely, it will be useful later to calculate the more general integral

$$I(n, D) = \int \frac{\mathrm{d}^D k}{(2\pi)^D} \left(\frac{1}{k^2 + i\epsilon} \right)^n \Theta(\Lambda' \leq |k| \leq \Lambda)\,, \tag{A.25}$$

defined in D space-time dimensions. As discussed above, we will perform a "Wick rotation"

$$I(n, D) = (-1)^n i \int \frac{\mathrm{d}^D k_E}{(2\pi)^D} \left(\frac{1}{k_E^2} \right)^n \Theta({\Lambda'}^2 \leq k_E^2 \leq \Lambda^2)\,. \tag{A.26}$$

The volume of the surface of the D-dimensional sphere is given by

$$\int \mathrm{d}\Omega_D = \frac{2\pi^{D/2}}{\Gamma(D/2)} \tag{A.27}$$

and therefore

$$I(n, D) = \frac{(-1)^n i}{(4\pi)^{D/2}\Gamma(D/2)} \int_{{\Lambda'}^2}^{\Lambda^2} \mathrm{d}k_E^2\, (k_E^2)^{D/2-1-n}\,, \tag{A.28}$$

which can be calculated explicitely

$$I(D/2, D) = \frac{(-1)^n i}{(4\pi)^{D/2}\Gamma(D/2)} \ln \frac{\Lambda^2}{{\Lambda'}^2} \tag{A.29a}$$

$$I(n, D)|_{n \neq D/2} = \frac{(-1)^n i}{(4\pi)^{D/2}\Gamma(D/2)} \frac{1}{D/2 - n} \left(\Lambda^{D-2n} - {\Lambda'}^{D-2n} \right)\,. \tag{A.29b}$$

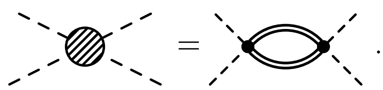
In the case of the tadpole diagram (A.24) we find finally

$$\int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{i}{k^2 + i\epsilon} \Theta(\Lambda' \leq |k| \leq \Lambda) = iI(1,4) = \frac{1}{(4\pi)^2} \left(\Lambda^2 - \Lambda'^2 \right) \tag{A.30}$$

for the new vertex. In the four point function

(A.31)

we need more vertices. First a modified four particle vertex



(A.32)

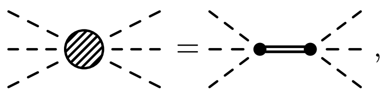
Under the assumption that the external momenta satisfy $|p| \ll \Lambda' < \Lambda$, we can do the loop integral in (A.32) easily to leading order

$$\begin{aligned} \frac{1}{2} \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{i\Theta(\Lambda' \leq |k| \leq \Lambda)}{k^2 + i\epsilon} \frac{i\Theta(\Lambda' \leq |p-k| \leq \Lambda)}{(p-k)^2 + i\epsilon} & \\ = -\frac{1}{2} \frac{i}{16\pi^2} \ln \frac{\Lambda^2}{\Lambda'^2} + O\left(\frac{p^2}{\Lambda^2} \right) . & \end{aligned} \tag{A.33}$$

The exact expression is more complicated, due to the finite integration limits, but we will not need its explicit form. More important is the observation that all integrals exist, since they are over a compact domain in Euclidean space, far from any singularities. Below, after motivating the physics, we shall see how to perform calculations more efficiently. Anyway, there are three such diagrams and the new vertex is

$$-\frac{3}{2} \frac{1}{16\pi^2} \ln \frac{\Lambda^2}{\Lambda'^2} \frac{g^2}{4!} \phi^4(x) . \tag{A.34}$$

There is also a new six particle vertex



(A.35)

that reads (again neglecting small terms $O(p^2/\Lambda^2)$):

$$\frac{1}{\Lambda^2} \frac{g^2}{6!} \phi^6(x) . \tag{A.36}$$

The general procedure consists in a stepwise reduction of the scale Λ'

$$\Lambda > \Lambda' > \Lambda'' > \Lambda''' > \ldots \tag{A.37}$$

where the Lagrangian is adjusted in each step such that the low energy observables below the current cut off scale is unchanged. If we allow arbitrary powers of the field ϕ and derivatives, the low energy physics can indeed be reproduced fully, as shown above.

A.2.4 Renormalization Group Flow

The most general interaction can be expanded

$$L(x) = \sum_i g_i O_i(x) \tag{A.38}$$

in an infinite series of local operators

$$O(x) = \left\{ \phi^2(x), \phi^4(x), (\partial\phi)^2(x), \phi^6(x), (\phi\partial\phi)^2(x), \ldots \right\} , \tag{A.39}$$

where an infinite sum will not necessarily correspond to a local interaction. A discrete renormalization group transformation of the scale dependent coupling constants

$$g_i(\Lambda) \to g_i(\Lambda') = \sum_j \Gamma_{ij}(\Lambda', \Lambda) g_j(\Lambda) \tag{A.40}$$

is defined by the requirement that the low energy physics remains unchanged.

However, discrete transformations like (A.40) are technically hard to handle. It is much more transparent to consider a continuous transformation with infinitesimal generators instead. Then the RGE is a system of coupled non-linear ordinary differential equations

$$\Lambda'^2 \frac{\mathrm{d}g_i(\Lambda')}{\mathrm{d}\Lambda'^2} = \sum_j \gamma_{ij} g_j(\Lambda') \,. \tag{A.41}$$

As a concrete example, consider the evolution of the coefficients g_4 and g_6

$$L_{\mathrm{int}}(x) = \frac{g_4}{4!}\phi^4(x) + \frac{g_6}{6!}\phi^6(x) + \ldots \tag{A.42}$$

with a change of the cut-off scale Λ', as depicted on the left hand side of figure A.8. By construction, the low-energy physics does not change along the trajectories in figure A.8, as long as each coupling uses the *same* Λ'.

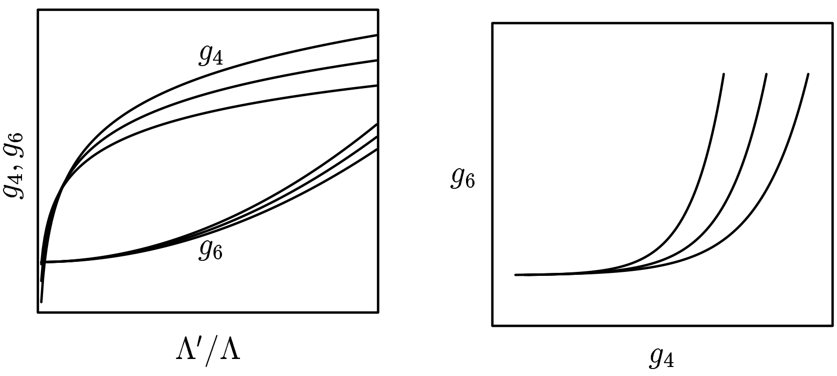


Figure A.8: *The evolution of the coefficients g_4 and g_6 as a function of a cut-off scale Λ' (left hand side) and after elimination of the unphysical cut-off scale Λ' (right hand side).*

The cut-off scale Λ' is therefore just *conventional* and determines what is considered to be part of the lowest order Lagrangian and what is to be calculated from it in higher orders. Thus the scale Λ' is redundant and can be eliminated from the physical parameter space. This physics is therefore not determined by a point in the parameter space, but by a whole *trajectory* instead.

Since Λ' is redundant, it is desirable to eliminate it completely. As long as $g_4(\Lambda')$ can be inverted (which is always possible in fixed orders of PT), the physics on the left hand side of figure A.8 can be recast as the right hand side of the same figure, where the dimensionful cut-off scale Λ' has be traded for the dimensionless coupling parameter g_4. This is an example of the celebrated phenomenon of *dimensional transmutation* [113], of which the relation of the strong coupling α_S and the QCD scale Λ_{QCD} is the most famous example.

So far, we have ignored the dependence on the upper cut-off scale Λ, which was assumed to be fixed and much larger than any physics scale. After the introduction of the RG trajectories, the scale Λ obtains a straightforward interpretation as the starting point of the trajectory. In this interpretation, the low energy physics will not depend on Λ if the RGE can be solved for all Λ. However, the existence of global solutions is not a trivial condition and there are several cases to consider.

In the scenario depicted on the left hand side of figure A.9, all trajectories extend to $\Lambda' \to \infty$. Therefore, the starting point of the evolution can be pushed back to $\Lambda \to \infty$, which means that the cut-off can be removed

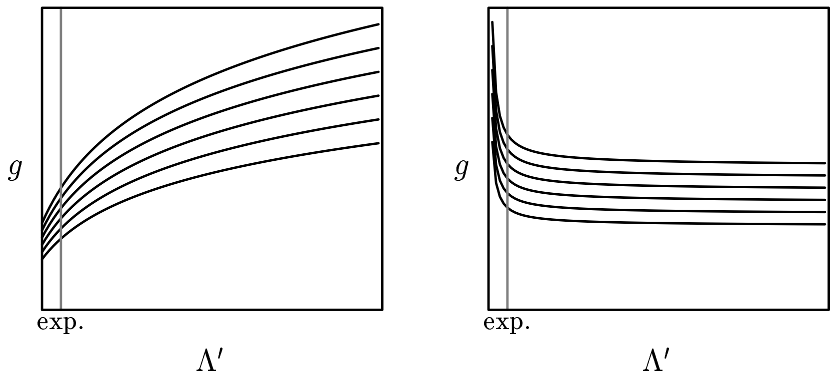


Figure A.9: *All trajectories on the left hand side extend to $\Lambda' \to \infty$ and a continuum limit exists. All trajectories on the right hand side even remain perturbative for $\Lambda' \to \infty$.*

altogether and a continuum limit exists. However, one must be careful that the trajectories might lead through non perturbative regions with $g \gg 1$ and that a perturbative determination of the evolution equation is unreliable in these regions.

Even more favorable is the scenario depicted on the right hand side of figure A.9, where all trajectories that start in the low energy regime, where couplings are determined experimentally, remain perturbative for all $\Lambda' \to \infty$. Perturbative calculations are reliable everywhere in this scenario. The most favorable special case of this scenario is *asymptotic freedom*, [109] in which the couplings vanish for $\Lambda' \to \infty$ and low orders of PT provide reliable predictions for high energy experiments.

In stark contrast to the previous two scenarios, no trajectory on the left hand side of figure A.10 can be extended from the low energy domain beyond the scale Λ_0. Thus the QFT that describes the present experiments can not be valid above Λ_0, unless non-perturbative effects change the trajectories significantly close to Λ_0. Such a scenario is not entirely bad news, because it predicts that something "interesting" will happen around Λ_0: at least strong interactions, maybe even "new physics."

An interesting combination of the scenarios in figures A.9 and A.10 is shown on the right hand side of figure A.10: some trajectories can be extended to $\Lambda' \to \infty$, while others are confined to $\Lambda' < \Lambda_0$. Typically, there will be a critical coupling $g_0(\Lambda')$ and trajectories with $g(\Lambda') < g_0(\Lambda')$ can be extended, while trajectories with $g(\Lambda') > g_0(\Lambda')$ can not. As a result, there will be upper limits on the couplings at any given scale, if the theory is in-

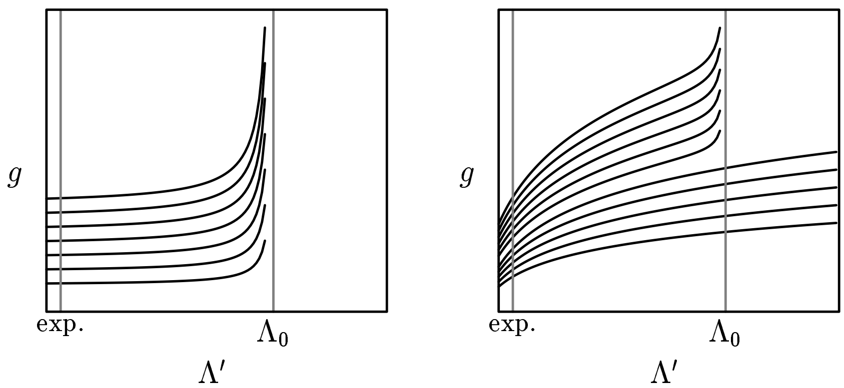


Figure A.10: *The low energy theory on the left hand side can never be valid above Λ_0. The low energy theory on the right hand side can be valid above Λ_0, but only for a restricted part of the parameter space.*

tended to be defined at arbitrary high scales. The most famous example for this phenomenon is provided by the upper limits on the Higgs self coupling resulting in an upper limit in the Higgs mass $m_H^2 = g/2 \cdot \langle \phi \rangle^2$ in the SM [114].

A.2.5 Relevant, Marginal Or Irrelevant?

While the graphical representation or the RG flow is very suggestive for a few couplings, the untruncated problem has to deal with *infinitely many* couplings. Therefore, it is not obvious that the intuition gained from the low-dimensional examples can be used in the infinite dimensional space of couplings. This is particularly problematic, because operators of higher and higher dimension produce more and more divergent contributions to Feynman diagrams and need even higher dimensional counterterms (see section A.1.2).

Fortunately, *"the first shall be last and the last shall be first"* and the higher dimensional operators turn out to be harmless from the point of view of the RG. If there was no interaction, the couplings would be constant. In this case, we can estimate matrix elements of the operators of dimension d by dimensional analysis as (in four space-time dimensions)

$$\left(\frac{|p|}{\Lambda}\right)^{d-4} \tag{A.43}$$

where Λ is the high energy scale at which the "bare" couplings are defined and d is the power counting dimension of the operator, counted according to

(A.9b) and (A.10). In (A.43), we can identify three different cases according to the value d

- $d < 4$: these operators are termed *relevant*, because their contributions are becoming more important at low energies, where experiments are performed.
- $d = 4$: these operators are termed *marginal*.
- $d > 4$: these operators are termed *irrelevant*, because their contributions are becoming less important at low energies, where experiments are performed.

Therefore, the initial conditions at high energies for higher dimensional ("non renormalizable") interactions play no important rôle in low energy physics.

Unless there are very strong interactions, relevant and irrelevant operators will retain their property, even if interactions are switched on. This can be seen from the evolution

$$\Lambda'^2 \frac{\mathrm{d}g_2(\Lambda')}{\mathrm{d}\Lambda'^2} = \Lambda'^2 \frac{1}{16\pi^2} g_4(\Lambda') + \ldots \tag{A.44a}$$

$$\Lambda'^2 \frac{\mathrm{d}g_4(\Lambda')}{\mathrm{d}\Lambda'^2} = \frac{3}{2} \frac{1}{16\pi^2} g_4^2(\Lambda') + \ldots \tag{A.44b}$$

$$\Lambda'^2 \frac{\mathrm{d}g_6(\Lambda')}{\mathrm{d}\Lambda'^2} = \frac{1}{\Lambda'^2} g_4^2(\Lambda') + \ldots \tag{A.44c}$$

of the low dimensional operators:

- $d = 2$ (i. e. g_2 or m^2): the rate of change of this coupling is of dimension 2: $O(\Lambda'^2)$. Unless there is significant *fine tuning* of the initial conditions, the value of the coupling will be of the order of the high energy starting point of the evolution: $g_2 = O(\Lambda^2) \gg \Lambda'^2$.
- $d = 6$ (i. e. g_6): the rate of change of the coupling is of dimension -2: $O(\Lambda'^{-2})$. Therefore the value of the initial condition at the high energy scale is indeed irrelevant with respect to the value induced by the evolution.

To avoid a possible misunderstanding induced by the technical term "irrelevant operator", it is important to stress that the contributions of irrelevant operators do *not* necessarily vanish in the infrared. Instead, the dependence of these contributions on their initial conditions at high energy vanishes, because the RG trajectories flow together in the infrared, as shown in figure A.11.

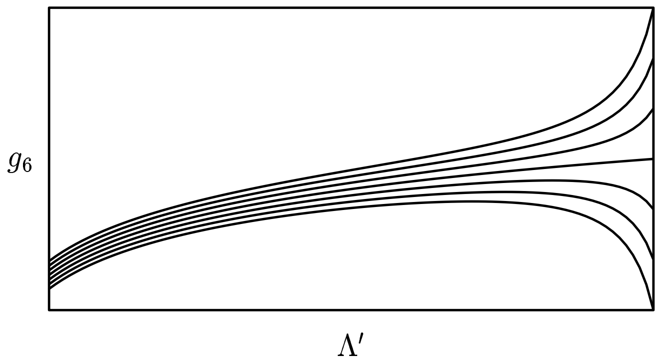


Figure A.11: *Contributions of irrelevant operators do* not *necessarily vanish in the infrared, but the dependence of these contributions on their initial conditions at high energy vanishes in the infrared.*

Thus the initial conditions for the irrelevant operators have no effect on the results of low energy experiments. In fact, the coefficients of the irrelevant operators can be set to zero at the high scale without changing the observable physics at lower energies. Therefore, the phenomena can be described by a renormalizable theory.

To avoid another possible misunderstanding, this observation does not "prove" that only a renormalizable QFT can describe the observed phenomena. It only states that there is always are renormalizable QFT that is indistinguishable from a non-renormalizable theory at low energies. Since renormalizable theories are technically more convenient and depend on less parameters, common sense and Occam's razor suggest to prefer the renormalizable QFT with only relevant or marginal interactions at the high scale over the others with the same infrared behavior.

The technical advantage of renormalizable theories lies in the fact that, for a given set of conserved quantum numbers, there is only a finite number of relevant or marginal operators, which is even quite small. This simplifies calculations and increases the predictive power of these theories because fewer free parameters have to be fitted in experiments.

In fact, from the point of view of the RG, the relevant ("super renormalizable") operators are more problematic than the non renormalizable ones. The most prominent examples are masses of light particles. Unless there is a symmetry that forces the mass to be small, most trajectories will lead to a value of the mass of the order of the high energy cut off scale, which is incompatible with the assumption that the particle is light. The selection of

Figure A.12: *Suppressed contributions to the β-function for g_4, which are $O(\Lambda^{-2})$.*

a trajectory that flows towards a light mass in the infrared requires unnatural fine tuning in this case.

Fermion masses can be protected by a chiral symmetry and need no fine tuning. The same applies to the Goldstone bosons of a spontaneous symmetry breaking. Goldstone's theorem protects their masses as well. Other scalars need to be related to fermions through a supersymmetry in order to be protected against a naturally large mass.

The marginal operators are the most interesting from the theoretical point of view. Without interaction, their contributions do not change with the scale. Therefore even weak interactions can make the marginal couplings grow strong or become weak asymptotically. From the RGE for g_4

$$\Lambda'^2 \frac{\mathrm{d}g_4(\Lambda')}{\mathrm{d}\Lambda'^2} = \beta(g_4(\Lambda')) \tag{A.45}$$

it can be seen that the behavior of the coupling depends on the sign of the β-function

- $\beta > 0$ (as in the scalar example considered above: $\beta = 3g_4^2/(32\pi^2)$): the coupling grows strong in the ultraviolet and becomes weak in the infrared,

- $\beta < 0$ (e.. g. QCD, see section A.2.9): the coupling grows strong in the infrared and becomes weak in the ultraviolet.

The solution of (A.45) for our scalar example is

$$g_4(\Lambda') = \frac{g_4(\Lambda)}{1 + \frac{3}{32\pi^2} g_4(\Lambda) \ln \frac{\Lambda^2}{\Lambda'^2}} \,. \tag{A.46}$$

Figure A.12 displays more contributions to the β-function for g_4, but all these contributions are $O(\Lambda^{-2})$ and can be ignored in leading order.

If the interactions are strong enough, relevant operators can become marginal or even irrelevant and vice versa. However, this is impossible in the perturbative domain and therefore outside our present scope. The non perturbative calculation of path integrals via the solution of a RGE is in principle possible, but technically very complicated. The known solutions require a truncation of the RGE and the systematic error introduced in this fashion is very hard to quantify.

A.2.6 Leading Logarithms

In addition to providing an explanation for the fortunate circumstance that nature seems to be described by a renormalizable QFT, the study of the RG flows provides a more reliable calculational tool than the direct perturbative evaluation of Feynman diagrams.

In theories with widely separated mass scales, PT suffers from large coefficients in the perturbative series that are caused by logarithmic integrals from one mass scale to the next. In contract, the coefficients of the RGE never contain more than one scale, which is determined by the slice of momentum space that is being integrated out. Thus these coefficients can not contain large logarithms and the theory remains perturbative as long as the couplings are weak. Observable large logarithms are recovered only in the final stage of the calculation, the solution of the RGE. Therefore, large unphysical logarithms that cancel in observables can never appear in intermediate steps.

A.2.7 Callan Symanzik Equations

The calculation of integrals over a small slice of momentum space like (A.25) or (A.33) is technically demanding, because the finite integration limits make it impossible to use many of the standard tricks of the trade, like the shifting of integration variables. In addition, fixed cut-offs manifestly violate gauge invariance, which makes their practical application in gauge theories [115] very awkward.

At this point, the relation of the RG flow to perturbative renormalization comes in handy. A change of the lower cut off scale Λ' corresponds to a change of the Lagrangian that keeps the physics unchanged. In perturbative renormalization, the renormalization scale μ plays a similar rôle as a scale that characterizes how much of the interaction is absorbed into the Lagrangian in the form of local counter terms. Therefore, it is tempting to identify Λ' with μ.

The analogy is not complete, however. The *hard* cut off Λ' is replaced by counter terms that correspond to a *soft* cut off, which is switched on gradually. This is the price to pay for gauge invariance and technical convenience. If we were to sum *all* orders of the perturbative series, both methods will give the same result. For a truncated series, however, differences of higher order will remain.

Consider a general n-point function

$$G^{(n)}(x_1, x_2, \dots, x_n; g, \mu) = \langle 0|\mathrm{T}\phi(x_1)\phi(x_2)\dots\phi(x_n)|0\rangle \,, \tag{A.47}$$

which is renormalized at μ, i. e. the invariants of the momenta in the renormalization conditions are equal to μ. A change of μ will not change the physics. Instead, the theory is shifted to another point on the *same* trajectory. Therefore, there is a new value g' for the renormalized coupling, such that only the unobservable normalization of the field operators is changed $\phi(x) \to 1/\sqrt{Z}\phi(x)$. Thus

$$G^{(n)}(x_1, x_2, \dots, x_n; g', \mu') = Z^{-n/2}(\mu, \mu')G^{(n)}(x_1, x_2, \dots, x_n; g, \mu) \,. \tag{A.48}$$

As above, it is more convenient to replace the discrete RG transformations by continuous transformations with infinitesimal generators

$$\mu' \frac{\mathrm{d}}{\mathrm{d}\mu'} \left(Z^{n/2}(\mu, \mu')G^{(n)}(x_1, \dots, x_n; g', \mu')\right) = 0 \tag{A.49}$$

where $\mathrm{d}/\mathrm{d}\mu'$ is a *total* derivative, that includes the change of the coupling g as well. Therefore, the Callan-Symanzik Equation (CSE) reads

$$\left(\mu\frac{\partial}{\partial\mu} + \beta(g)\frac{\partial}{\partial g} + n\gamma(g)\right) G^{(n)}(x_1, \dots; g, \mu) = 0 \,, \tag{A.50}$$

with

$$\beta(g) = \mu\frac{\mathrm{d}}{\mathrm{d}\mu}g(\mu) \tag{A.51a}$$

$$\gamma(g) = \frac{1}{2Z(\mu_0, \mu)}\mu\frac{\mathrm{d}}{\mathrm{d}\mu}Z(\mu_0, \mu) \,. \tag{A.51b}$$

The β-function is dimensionless. In a massless theory, it can therefore not depend on μ for dimensional reasons. In theories with masses, there are always renormalization schemes that retain this property [116]. In other schemes, β can depend on ratios of μ and masses.

However, it is *not* obvious that γ does not depend on μ_0 through the ratio μ/μ_0. In renormalizable theories, the limit $\mu_0 \to \infty$ exists and γ can not depend on μ_0. As we have explained in section A.2.5 above, we can always chose a trajectory that corresponds to a renormalizable theory and drop the dependence of γ on μ_0.

The actual calculation of the coefficient functions β and γ proceeds by calculating a sufficiently large set of renormalized Green's functions and finding functions β and γ such that (A.50) is satisfied. The non-trivial statement of the Callan-Symanzik Equation (CSE) is that one universal function β for each coupling and one universal function γ for each field suffices to satisfy (A.50) for *all* Green's functions.

If there is more than one coupling and more than one field, the generalization is obvious

$$\left(\mu\frac{\partial}{\partial\mu} + \sum_{i=1}^{k}\beta_i(g)\frac{\partial}{\partial g_i} + \sum_{j=1}^{m} n_j\gamma_j(g)\right) G^{(n_1,\ldots,n_m)}(x_1,\ldots;g_1,\ldots,g_k,\mu) = 0\,. \tag{A.52}$$

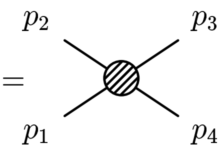
For example, the QED vertex functions obeys the CSE

$$\left(\mu\frac{\partial}{\partial\mu} + \beta(e)\frac{\partial}{\partial e} + 2\gamma_\psi(e) + \gamma_A(e)\right)\left\langle 0\middle|\mathrm{T}\psi(x_1)\bar\psi(x_2)A_\nu(x_3)\middle|0\right\rangle = 0\,. \tag{A.53}$$

A.2.8 Running Couplings and Anomalous Dimensions

As a concrete example for solving the CSE, consider a four-point function

$$G^{(4)}(P;g,\mu) =$$



$$= \mathrm{F.T.}\,\langle 0|\mathrm{T}\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)|0\rangle \tag{A.54}$$

in the deep Euclidean domain (this does not correspond to a physical amplitude, but can be part of one):

$$p_i^2 = -P^2 \tag{A.55a}$$

$$p_ip_j = 0\,. \tag{A.55b}$$

Lowest order PT gives for the connected (but not amputated) part

$$G^{(4)}(P;g,\mu) = \left(\frac{-i}{P^2}\right)^4(-ig) + O(g^2)\,. \tag{A.56}$$

Ignoring corrections from mass terms, dimensional analysis predicts that to all orders in PT

$$G^{(4)}(P;g,\mu) = \left(\frac{-i}{P^2}\right)^4\hat G^{(4)}(P/\mu;g)\,, \tag{A.57}$$

i. e. that $G^{(4)}$ is a homogenous function and solves the differential equation

$$\mu\frac{\partial}{\partial\mu}G^{(4)}(P;g,\mu) = -\left(8 + P\frac{\partial}{\partial P}\right)G^{(4)}(P;g,\mu)\,. \tag{A.58}$$

Plugging (A.58) into the CSE (A.50), we can trade the dependence of the renormalization scale for the momentum dependence and obtain the Renormalization Group Equation (RGE) for the four point function

$$\left(P\frac{\partial}{\partial P} - \beta(g)\frac{\partial}{\partial g} + 8 - 4\gamma(g)\right)G^{(4)}(P;g,\mu) = 0\,. \tag{A.59}$$

The RGE can be solved with standard techniques for the integration of a Partial Differential Equation (PDE). The introduction of a *running coupling constant* $\bar{g}(P;g)$

$$P\frac{\mathrm{d}}{\mathrm{d}P}\bar{g}(P;g) = \beta(\bar{g}(P;g)) \tag{A.60a}$$

$$g(\mu;g) = g \tag{A.60b}$$

absorbs the differential operator

$$\left(P\frac{\partial}{\partial P} - \beta(g)\frac{\partial}{\partial g}\right)\bar{g}(P;g) = 0 \tag{A.61}$$

and the general solution can be written as

$$G^{(4)}(P;g,\mu) = \left(\frac{-i}{P^2}\right)^4 \mathcal{G}^{(4)}(\bar{g}(P;g)) \cdot \exp\left(4\int\limits_{g}^{\bar{g}(P;g)} \mathrm{d}g'\,\frac{\gamma(g')}{\beta(g')}\right)\,, \tag{A.62}$$

where $\mathcal{G}^{(4)}$ is an *arbitrary* function of one variable, which is *not* determined by the RG. However, *matching* the solution (A.62) of the RGE to the perturbative result (A.56) at $P = \mu$ gives

$$\mathcal{G}^{(4)}(\bar{g}) = -i\bar{g} + O(\bar{g}^2)\,. \tag{A.63}$$

In general, we can identify two elements in the solution of the RGE. One part of the solution is given by the perturbative result at a given order with the coupling constant replaced by the running coupling at the appropriate scale. In addition, there is an exponential factor modifying the scaling behavior of each field in the n-point function under consideration. Since the γ appear in the numerator of the exponent, they are usually called *anomalous dimensions*.

From the β-function for ϕ^4

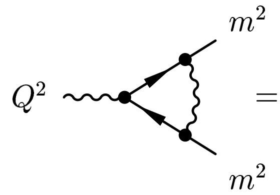
$$\beta(g) = \frac{3}{16\pi^2} g^2 + O(g^3) \tag{A.64}$$

follows the running coupling

$$\bar{g}(P; g) = \frac{g}{1 - \frac{3}{16\pi^2} g \ln \frac{P}{\mu}} \, . \tag{A.65}$$

The solution (A.62) has the nice feature that *all* large logarithms $\ln(P^2/\mu^2)$ are collected in two universal places, that are common to all Green's functions: the running coupling $\bar{g}(P; g)$ and the exponents of the scale factors. Thus the large logarithms have been resummed successfully in universal quantities and PT is reliable, as long as $\bar{g}(P; g)$ is small.

Unfortunately, this procedure does not work as straightforwardly for all Green's functions. The RGE is much harder to solve if not all external momenta scale uniformly. If there is more than one relevant mass scale, the coefficient functions themselves can contain large logarithms. The classical example for large logarithms that can not be obtained from the RG are the Sudakov double logarithms [117] in exclusive scattering processes



$$= -\frac{\alpha}{2\pi} \ln^2 \left(\frac{Q^2}{m^2} \right) \tag{A.66}$$

A.2.9 Asymptotic Freedom

As discussed earlier, the behavior of the β-function in the vicinity of $g = 0$ determines the properties of PT. For a single coupling, the β-function must vanish for $g = 0$, because without interactions, there is no mechanism that could induce running of the coupling.

Figures A.13 and A.14 display the four possible different scenarios. On the left hand side of figure A.13, we have $\beta > 0$ everywhere and the running coupling grows in the ultraviolet without limits. However, PT is reliable in the infrared.

On the right hand side of figure A.13, we have $\beta < 0$ everywhere and the running coupling grows in the infrared without limits. PT is reliable in the ultraviolet, but problematic in the infrared. The theory of strong interactions, QCD, has this property [109], called *asymptotic freedom* and it is known that non abelian gauge theories are the only class of theories in four space-time dimensions that can exhibit asymptotic freedom.

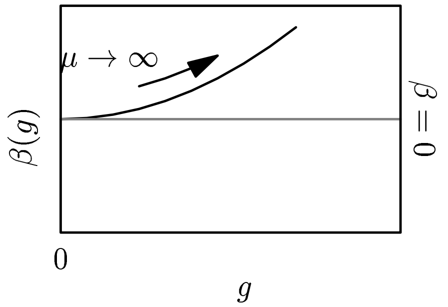


Figure A.13: *β-functions for infrared free (left) and ultraviolet free (right) theories. The arrow denotes the direction of the evolution for a rising scale.*

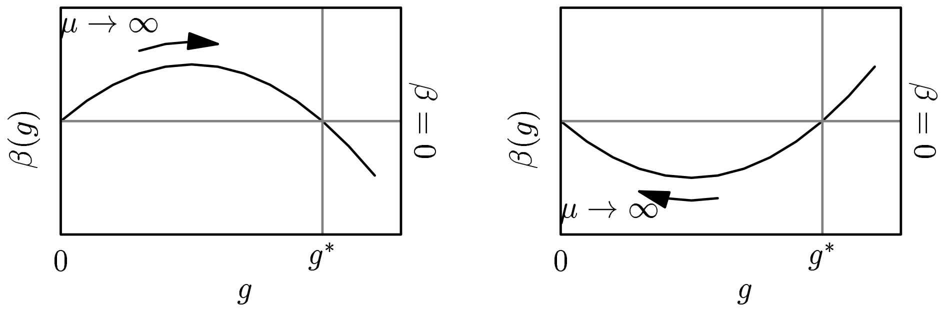


Figure A.14: *β-functions with ultraviolet (left) and infrared (right) stable fixed points. The arrow denotes the direction of the evolution for a rising scale.*

In the left hand side of figure A.14, the coupling grows in the ultraviolet, but the β-function makes a turn and the coupling becomes constant after running into an *ultraviolet stable fixed point* at g^*, because $\beta(g^*) = 0$. The opposite phenomenon appears in the right hand side of figure A.14, where g^* is an *infrared stable fixed point.*

A.2.10 Dimensional Regularization

As mentioned above, the RG functions $\beta(g)$ and $\gamma(g)$ are universal. They are determined by calculating $\mu\partial G^{(n)}/\partial\mu$ for a sufficient set of $G^{(n)}$ and solving the CSE for $\beta(g)$ and $\gamma(g)$.

A typical integral appearing in these equations is

$$\hat{I}_{n,m}(D, M^2) = \mu^{D-4} \int \frac{\mathrm{d}^D k}{(2\pi)^D} \frac{(k^2)^n}{(k^2 - M^2 + i\epsilon)^m}\,, \tag{A.67}$$

where the factor μ^{4-D} has been added in order to make the mass dimension of $\hat{I}$ independent of the dimension of space time. As long as $m - n - D/2$ and $D/2 + n$ are not equal to negative whole numbers, the integral can be evaluated

$$\hat{I}_{n,m}(D, M^2) = \frac{(-1)^{n+m} i}{16\pi^2} \left(M^2\right)^{2+n-m} \cdot \left(\frac{M^2}{16\pi^2\mu^2}\right)^{D/2-2} \frac{\Gamma(m-n-D/2)\Gamma(D/2+n)}{\Gamma(m)\Gamma(D/2)} \tag{A.68}$$

and the logarithmic divergency of $\hat{I}_{0,2}(4, M^2)$ shows up as the pole of the Euler Γ-function. This divergency is regularized by moving away from $D = 4$

$$\hat{I}_{0,2}(4 - 2\epsilon, M^2) = \frac{i}{16\pi^2} \left(\frac{M^2}{16\pi^2\mu^2}\right)^{-\epsilon} \Gamma(\epsilon)\,. \tag{A.69}$$

$\Gamma(\epsilon)$ can be expanded in a Laurent series, that starts with a single pole

$$\hat{I}_{0,2}(4 - 2\epsilon, M^2) = \frac{i}{16\pi^2} \left(\frac{1}{\epsilon} + \ln\frac{\mu^2}{M^2} + 2\ln 4\pi - \gamma_E\right)\,. \tag{A.70}$$

In a massless theory, the dependence on μ must be identical to the dependence on the renormalization scale and the expansion of the Γ-function shows that it suffices to determine the residues of the poles in $\epsilon = 2 - D/2$. Even in massive theories, there are renormalization schemes that retain this convenient property [116].

This procedure of *dimensional regularization* [108] and *minimal subtraction* of the poles in $1/(D-4)$ is particularly useful in gauge theories, because gauge invariance is not violated by the regularization and no additional counter terms are needed in order to maintain gauge invariance. Since gauge invariance softens divergencies, dimensional regularization avoids the overestimation of radiative corrections induced by a naive application of regularization procedures that violate gauge invariance.

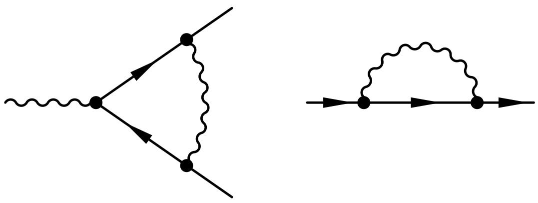


Figure A.15: *The contributions of the vertex function and the self energy to the β-function in QED cancel.*

A.2.11 Gauge Theories

In QED, we find for the lowest order RG functions

$$\beta(e) = \frac{e^3}{12\pi^2} \tag{A.71a}$$

$$\gamma_\psi(e) = \frac{e^2}{16\pi^2} \tag{A.71b}$$

$$\gamma_A(e) = \frac{e^2}{12\pi^2} \, . \tag{A.71c}$$

There is a peculiarity of QED: the Ward identity enforces that the contributions of the vertex function and the self energy (see figure A.15) to the β-function cancel to all orders and the β-function can be calculated from the vacuum polarization (figure A.3) alone.

From (A.71a), we see that $\beta_{\mathrm{QED}} > 0$ and that QED is not asymptotically free. Higher order calculations have not revealed an ultraviolet fixed point either.

A more complicated calculation in a non-abelian SU(N_C) gauge theory with N_f quarks has the result [109]

$$\beta(g) = \frac{e^3}{16\pi^2}\left(C_F\frac{N_f}{2} - \frac{11}{3}N_C\right) \, , \tag{A.72}$$

where

$$C_F = \frac{N_C^2 - 1}{2N_C} \, . \tag{A.73}$$

In QCD (i. e. $N_C = 3$), this means

$$\beta_{\mathrm{QCD}}(g) = \frac{e^3}{16\pi^2}\left(\frac{2}{3}N_f - 11\right) \tag{A.74}$$

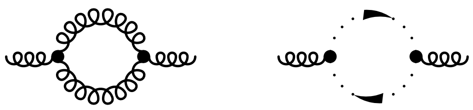


Figure A.16: *Non abelian contributions to the vacuum polarization: gluon and ghost loops.*

and the theory is indeed asymptotically free for $N_f < 33/2$. As a result, the running coupling of QCD falls with rising energy:

$$\alpha_s(Q^2) = \frac{4\pi}{\beta_0 \ln(Q^2/\Lambda^2_{\mathrm{QCD}})} \tag{A.75}$$

where $0.15\,\mathrm{GeV} < \Lambda_{\mathrm{QCD}} < 0.25\,\mathrm{GeV}$ and $\beta_0 = 11 - 2N_f/3$.

The most obvious difference to an abelian gauge theory like QED is the existence of the genuinely non-abelian contributions to the vacuum polarization displayed in figure A.16. These diagrams do indeed give a negative contribution to the β-function.

Intuitively, the negative contribution from the gluon and ghost loops gives a fair description of the interplay of the positive quark contribution and the negative non-abelian contributions. Strictly speaking, however this explanation is not correct, because the cancellations of vertices and self energies familiar from QED are not operative and more complicated diagrams also give contributions to the β-function.

A.3 Effective Field Theory

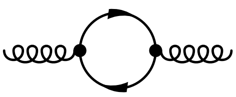
Ein Bild erhält ja gerade dadurch seine
Brauchbarkeit, daß es nicht *die Sache ist.*
[EGON FRIEDELL: *Kulturgeschichte Griechenlands*]

A.3.1 Flavor Thresholds and Decoupling

So far, all effects of particle masses have been neglected, in order to highlight the simplicity of the RG approach. This is often a very good approximation. For example at LEP2, fermion masses can be neglected in four fermion production, unless one of the final state particles is an electron and a t-channel singularity has to be regularized. The massless approximation is a good approximation, because the LEP2 center of mass energy is far from all flavor thresholds: $m_b \ll \sqrt{s_{\mathrm{LEP2}}} < m_t$. Top quarks can not be produced and m_b/M_W is a small number.

Close to a threshold, the RG loses some of its appeal, because the value of the threshold introduces a second scale that enters the calculation of the RG functions and makes calculations more complicated. However, Effective Field Theory (EFT) [118] provides a systematic way of dealing with flavor thresholds in a RG calculation.

Recall the β-function of a $\mathrm{SU}(N_C)$ gauge theory (A.72). The quarks loops in the gluon vacuum polarization induce a dependence on the number of quark flavors

p  $\propto \frac{e^3}{16\pi^2} C_F \frac{N_f}{2}$ (A.76)

and in a concrete calculation we have to answer the question for the value of N_f. This question can not be answered by counting the quarks in table 2.1 alone. To see this, consider the typical case $m_b^2 \ll -p^2 \ll m_t$ and $m_b^2 \ll \mu^2 \ll m_t$. Then the contribution of a top quark loop in (A.76) is

$$\Pi^R(p^2) = O(\frac{p^2}{m_t^2}, \frac{\mu^2}{m_t^2}) \tag{A.77}$$

and is very small below the threshold.

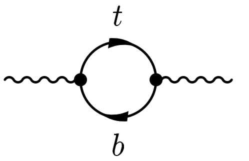
The purpose of the RG is to systematically resum the large logarithms in the vacuum polarization and other diagrams, but we find that there are no large logarithms below the threshold, as long as the renormalization scale is chosen sensibly. Therefore, the top flavor must not be included in (A.76) below threshold and $N_f = 5$ is the appropriate value between the bottom and top thresholds.

The fact that a heavy particle does not contribute to RG functions below its threshold is non-trivial. In the unrenormalized version of (A.76), each flavor contributes equally to the high energy tail of the loop integral, as long as the cut off is significantly above the threshold of this flavor. Only after renormalization, it can be seen that each flavor also contributes equally to the counter term and that the net contribution of a heavy flavor vanishes, if the Green's functions are renormalized below the threshold of this flavor.

Since the RG never deals with the loop integrals over all of momentum space, the cancellations described in the previous paragraph take place automatically. The β-function describes the rate of change of the running coupling and if only active flavors contribute to this change, the cancellation of contributions from heavy flavors and their counter terms is already manifest. This argument, which has been made heuristically using (A.76) for the β-function, can be made rigorous in the *decoupling theorem* [119]; also for the anomalous dimensions.

However, there is one circumstance under which the effects of a heavy particle can be felt in the RG flow *below* its threshold. This the case if the large mass violates a symmetry of the theory. The most celebrated example of this effect is the breaking of the $\mathrm{SU}(2)_c$ custodial symmetry [30] of the SM through the heavy top mass and the corresponding contribution to the ρ-parameter. In a perturbative calculation with masses, there is a contribution to ρ that is proportional to the quadratic mass splitting in the third generation

$$\Pi_{W^+W^-}(q^2) =$$



$$\propto \frac{\alpha}{\pi}\frac{m_t^2 - m_b^2}{M_W^2} + \cdots \tag{A.78}$$

If every $\mathrm{SU}(2)_L$ doublet was degenerate, the decoupling theorem would apply and each heavy doublet would decouple from the low energy phenomena. The resulting low energy physics could be described by an EFT built from $\mathrm{SU}(2)_c$-invariant interactions. If, as in the third generation of the SM, there is a strong lifting of the degeneracy, there will be an effect on the low energy couplings that manifests itself as a ρ-parameter that is different from unity. Then the low energy EFT below the top threshold has to contain $\mathrm{SU}(2)_c$-violating operators. Below the bottom threshold, there are no new $\mathrm{SU}(2)_c$ violating contributions from the fermion sector, but the effects accumulated between m_t and m_b remain. If there would be no reason to expect a custodial symmetry or $\rho = 1$, the low energy EFT would contain $\mathrm{SU}(2)_c$-violating operators from the outset, of course.

The general result of the decoupling theorem is that, unless symmetries are violated, heavy particles can be removed from a theory, if only phenomena below the energy scale given by the particle's mass are to be calculated. The only residual effect of the heavy particle remains in a renormalization of coupling constants and fields.

A.3.2 Integrating Out or Matching and Running

The decoupling of heavy flavors discussed in the previous section provides a systematic calculational procedure for theories with more than one mass scale. For example, in order the calculate the strong coupling at low energies from an experimental measurement at a 1 TeV LC, one proceeds as follows:

1. start with the measured value g_0 at $\mu_0 = 1\,\mathrm{TeV} \gg m_t$: $\bar{g}(\mu_0) = g_0$
2. solve the massless RGE $\bar{g}(\mu)$ with the β-function for six active flavors from $\mu = \mu_0$ down to $\mu = m_t$

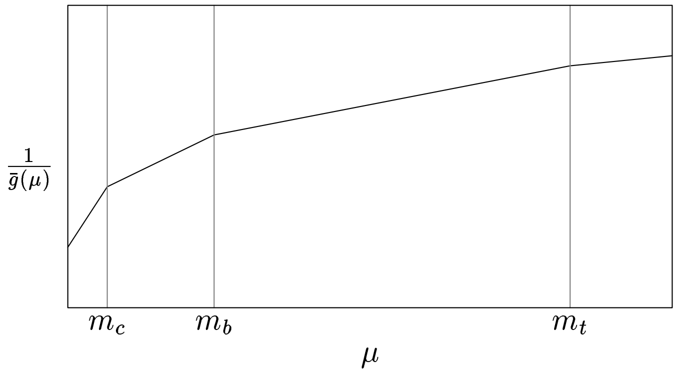


Figure A.17: *Running couplings with thresholds. In an asymptotically free gauge theory, the running is slowed down at each new flavor threshold.*

3. start a second evolution step at $\mu = m_t$, using the result of the first step as initial value
4. solve the massless RGE $\bar{g}(\mu)$ with the β-function for five active flavors from $\mu = m_t$ down to $\mu = m_b$.
5. start a third evolution step at $\mu = m_b$, ...
6. and so on

This procedure is known *matching and running* and is illustrated in figure A.17. The matching at each threshold is continuous, but not differentiable, because the rate of change (i. e. the β-function) changes discontinuously at each threshold.

The approximation of the flavor thresholds by step functions is not optimal, of course. However, for many observables, the detailed features of the threshold are much less important than the large logarithms between thresholds and the matching and running procedure provides a systematic way of resumming these logarithms. In any case, the calculations can be improved systematically by including higher orders (NLO, NNLO) in the loop expansion, as well as power corrections.

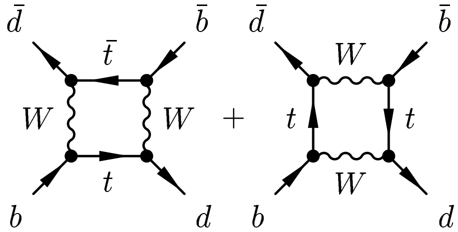
The most famous example (already mentioned in section 1.2) for a threshold effect is given by the running coupling of the three SM gauge groups. The experimentally determined couplings do *not* meet [11] in one point as would be a necessary condition of a GUT. However, matching the evolution of the couplings in the SM to the evolution of the couplings in the MSSM at the threshold provided by the supersymmetry breaking scale $\mu \approx 1\,\mathrm{TeV}$ causes the three couplings to meet [11]. This may be an accident, but it is the most suggestive phenomenological evidence for supersymmetry so far.

A.3.3 Important Irrelevant Interactions

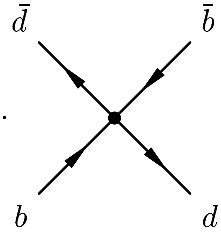
We have argued in section A.2.5, that operators of dimension greater than four are irrelevant and that the influence of their initial value at a high matching scale on the physics at much lower energies can be ignored.

This argument can not be applied in two important cases. If the irrelevant interaction violates a symmetry that is respected by all relevant and marginal operators, the effects of the irrelevant interactions can be observable, even if they are small in absolute terms. A different problem arises if the hierarchy of scales is not large enough and the lever arm of the RG is therefore not large enough to suppress irrelevant operators when running from one threshold to the next.

A typical example for the latter case is provided by the thresholds for the top quark and the $W^\pm$ bosons. The power corrections $M_W/m_t \approx 0.45$ can not be neglected relative to the logarithmic terms $|\ln(M_W/m_t)| \approx 0.8$. The solution in cases like $B^0\bar{B}^0$-mixing is to decouple the top quark and W boson *simultaneously*



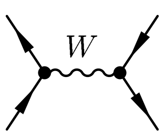
$$= \Phi\left(\frac{M_W^2}{m_t^2}\right) \cdot$$



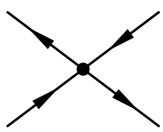
$$, \qquad \text{(A.79)}$$

where the function Φ is calculated from the massive loop [120]. Since the scales in the loop are rather close, no large logarithms can develop and the loop expansion for Φ is well behaved.

The $|\Delta B| = 2$ operator discussed in the previous paragraph is also an example for first case, but the more common example is provided by tree level charged current interactions. Below M_W, the exchange of a W can be approximated by a local interaction



$$=$$



$$+ O\left(\frac{p^2}{M_W^2}\right) \qquad \text{(A.80)}$$

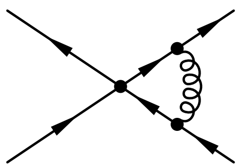
of dimension six

$$L_F = \frac{G_F}{\sqrt{2}} \bar{\psi}(1-\gamma_5)\gamma_\mu\psi\, \bar{\psi}(1-\gamma_5)\gamma^\mu\psi \qquad \text{(A.81)}$$

which is irrelevant. However, the Fermi interaction L_F can not be ignored, because it includes flavor changing terms that are absent in all relevant or marginal operators provided by the electromagnetic and strong interaction.

The SM $W^{\pm}$-exchange is matched to the local Fermi interaction (A.81) at $\mu = M_W$ like (A.80). Below the W-mass, all calculations are performed using the local Fermi interaction. The irrelevant operator (A.81) is not renormalizable in the traditional perturbative sense, because it requires counter terms of higher dimensions. In the EFT, this is no problem, however, because these counter terms are all determined by matching conditions at M_W. Consequently, there is no loss of predictive power.

The radiative corrections can be summed below M_W by solving a RGE that includes the anomalous dimension of the irrelevant operator that is derived from diagrams like

 $+ \ldots$

A.3.4 Mass Independent Renormalization

A useful intuitive picture for the matching and running procedure is that more and more physics from higher scales is *integrated out* as the renormalization scale is lowered. The physics is not lost in this process, but is condensed from Feynman diagrams into the short distance interaction vertices. One can imagine this as if all momentum integrations would be carried out from above down to the renormalization scale and that the remaining low energy momentum integrations would be cut off at the renormalization scale.

This intuitive picture is, however, not entirely accurate. As mentioned above, a fixed momentum space cut off is technically inconvenient, because of complicated integrals. But more important is that the fixed cut off is physically dangerous (see e. g. [37]), because gauge symmetries are violated by cut-offs. Therefore the dividing line between the short distance physics, which is absorbed in effective local interactions, and the long distance physics is gauge dependent and thus unphysical. Typically, the unphysical short distance coefficients will be spuriously large in magnitude, but their contributions will cancel in physical observables against unphysical long distance contributions. In order to exploit the cancellation of unphysical contributions as soon as possible, it is preferable to extend all loop integrals over the whole momentum space and to replace the hard cut offs by counter terms.

The price to pay is small, because the resulting divergencies can be handled with traditional perturbative renormalization theory. Calculations in mass independent scheme like minimal subtraction are convenient and do respect gauge invariance. The physically important threshold effects are handled by matching.

A.3.5 Naive Dimensional Analysis

If the theory at the high energy scale is known, EFT is but a very powerful calculational tool. It allows to use the most appropriate degrees of freedom at each scale by matching different theories at thresholds. Furthermore, large radiative corrections can be summed systematically through the renormalization group.

In such a scenario, the short distance coefficients are calculated by matching and running from the top to the bottom. The opposite case is when low energy data is fitted to an arbitrary effective Lagrangian and this Lagrangian is evolved from bottom to top until it becomes inconsistent, as shown in figure A.10. In this case, reliable estimates of the scale of "new physics" can be made, but the "new physics" above the threshold can rarely be determined from the matching conditions at the threshold.

Nevertheless, we can use EFT to estimate the likely size of contributions from "new physics" to observables at future experiments using Naive Dimensional Analysis (NDA) [110, 121, 23]. We know from general RG arguments, that the relevant operators will have larger coefficients at low energy than the irrelevant operators of the same symmetry. That this statement remains true after renormalization was first pointed out in [110]. Unfortunately these arguments alone give no indication of the absolute magnitude of the coefficients.

From (A.30) and (A.33) we see that typical loop integrals that contain a single scale can be estimated by a factor of

$$\frac{1}{16\pi^2} \approx 6.3 \cdot 10^{-3}\,, \tag{A.82}$$

which is multiplied by a power of this scale as dictated by simple dimensional analysis. A systematic analysis [110, 121] reveals that the small number (A.82) creates a hierarchy that can be used classify terms.

One possible form[4] of the lowest order effective Lagrangian in the momentum expansion [110] for a non-linearly realized broken abelian[5] symmetry is

$$L = f^2 \operatorname{tr}\left((D_\mu U)^2 D^\mu U\right) \tag{A.83a}$$

where

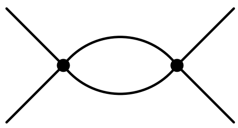
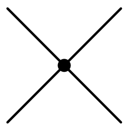
$$U(x) = e^{i\phi(x)/f} \tag{A.83b}$$

and f is a scale introduced by the symmetry breaking. It corresponds to the vacuum expectation value of a Higgs field in a linear realization of the symmetry breaking.

Loops renormalize the operators of a given order as well as higher orders in the momentum expansion. For example, a loop with two lowest order terms

$$\frac{1}{f^2}\left(\phi(x)\overleftrightarrow{\partial_\mu}\phi(x)\right)\left(\phi(x)\overleftrightarrow{\partial^\mu}\phi(x)\right) \tag{A.84}$$

renormalized at a scale μ^2

 $= \dfrac{1}{16\pi^2}\dfrac{P^2}{f^2}\ln\left(\dfrac{\mu^2}{P^2}\right)\cdot$  $+\ldots\,,$ (A.85)

where P^2 is an invariant built from the external momenta, requires a counterterm from the next order of the momentum expansion

$$\frac{1}{\Lambda^2}\frac{1}{f^2}\partial_\nu\phi(x)\partial_\mu\phi(x)\partial^\nu\phi(x)\partial^\mu\phi(x)\,, \tag{A.86}$$

where Λ is the scale characterizing the momentum expansion. A priori, there is *no* relation between the scales f and Λ for symmetry breaking and momentum expansion, respectively. However, unless there is some unnatural fine-tuning of parameters at work, the rate of change of the couplings in the RG running from (A.85) *should* be of the same order as the coefficients of the counterterm (A.86). Equating coefficients gives

$$\frac{1}{16\pi^2}\frac{1}{f^2} \approx \frac{1}{\Lambda^2} \tag{A.87}$$

[4] There are many equivalent forms [25], but *all* involve the field $\phi(x)$ only in the combination $\phi(x)/f$.

[5] In the general non-abelian case, the formulae are more complicated, because the factor group manifold G/H has to be parametrized [25]. These geometrical complications have no impact on the power counting argument.

or the celebrated formula

$$\Lambda \approx 4\pi f \,. \tag{A.88}$$

From another perspective, the RG will evolve towards (A.88) in the infrared, unless a symmetry that has not been made manifest in the effective Lagrangian forces the RG flow in a particular direction. A physical threshold between f and Λ would require matching and provides the only way around (A.88) in an EFT. Such a physical threshold will however be observable independently.

The heuristic argument for the lowest order in the momentum expansion for scalar fields can be extended systematically to all orders in the loop and momentum expansions and to fermions and vector fields [121]. The result is that a natural interaction involving scalars ϕ, fermions ψ, derivatives ∂ and gauge bosons A has the size

$$L^{(N_\phi,N_\psi,N_A,N_\partial)} = f^2\Lambda^2 \left(\frac{\phi}{f}\right)^{N_\phi} \left(\frac{\psi}{f\sqrt{\Lambda}}\right)^{N_\psi} \left(\frac{eA}{\Lambda}\right)^{N_A} \left(\frac{\partial}{\Lambda}\right)^{N_\partial} \tag{A.89}$$

for any quartet $(N_\phi, N_\psi, N_A, N_\partial)$ of natural numbers. The power counting rules assume that the vector boson A is indeed a gauge boson, such that the two parts of the covariant derivative $\partial - ieA$ scale identically. The rules would be much more complicated for non-gauge bosons and their longitudinal polarization states.

The important result of NDA is that the natural strengths (A.89) of interactions can be estimated, even in a strongly interacting EFT. The assumption that couplings have their natural strengths is nothing more than an assumption, of course. Nevertheless, the predictions from NDA work for low energy strong interactions and NDA can be used as a guide while searching for good places for hunting "new physics." Higher dimensional operators are most likely suppressed by a factor of $1/(16\pi^2)$ and proposals for experiments should take this into account.

Bibliography

[1] P. Zerwas, *Physics with e^+e^- Linear Colliders*, Technical Report No. DESY 94-001-REV, DESY (1996).

[2] H. Murayama and M. E. Peskin, Annu. Rev. Nucl. Part. Sci. **46**, 533 (1997).

[3] E. Accomando *et al.*, Phys. Rep. **299**, 1 (1998).

[4] S. Glashow, Nucl. Phys. **22**, 579 (1961); S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967); A. Salam, in *Elementary Particle Theory*, edited by W. Svartholm (Almquist and Wiksell, Stockholm, 1968), p. 367.

[5] D. Bardin *et al.*, in *Reports of the working group on precision calculations for the Z resonance*, CERN, edited by D. Bardin, W. Hollik, and G. Passarino (CERN, Geneva, 1997), No. CERN 95-03A.

[6] LEP, ALEPH, DELPHI, L3, and OPAL Collaborations, LEP Electroweak Working Group and SLD Heavy Flavor and Electroweak Groups, *A Combination of Preliminary Electroweak Measurements and Constraints On the Standard Model*, Technical Report No. CERN-EP-99-015, CERN, Geneva (1999).

[7] S. Eidelman and F. Jegerlehner, Z. Phys. C **67**, 585 (1995); M. Davier and A. Höcker, Phys. Lett. **435**, 427 (1998); J. Kühn and M. Steinhauser, Phys. Lett. **B437**, 425 (1998).

[8] Particle Data Group, Phys. Rev. **D54**, 1 (1996).

[9] Particle Data Group, Eur. Phys. J. **C3**, 1 (1998).

[10] S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. **D2**, 1285 (1970).

[11] U. Amaldi, W. de Boer, and H. Fürstenau, Phys. Lett. **B260**, 447 (1991).

[12] H. Georgi, in *Perspectives on Higgs Physics*, Vol. 13 of *Advanced Series on Directions in High Energy Physics*, edited by G. L. Kane (World Scientific, Singapore – New Jersey – London – Hong Kong, 1993), p. 337.

[13] ALEPH, DELPHI, L3, OPAL and D0 TGC Combination Group, *A Combination of Preliminary Measurements of Triple Gauge Boson Coupling Parameters Measured by the LEP and D0 Experiments*, Technical Report No. LEPEWWG/TGC/98-02, CERN and FNAL, Geneva and Batavia, IL (1998).

[14] C. Arzt, M. B. Einhorn, and J. Wudka, Nucl. Phys. **B433**, 41 (1995).

[15] J. Ellison and J. Wudka, Annu. Rev. Nucl. Part. Sci. **48**, 33 (1998).

[16] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, Phys. Lett. **B429**, 263 (1998); N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, Phys. Rev. **D59**, 086004 (1999).

[17] L. Hall and C. Kolda, *Electroweak Symmetry Breaking and Large Extra Dimensions*, Technical Report No. UCB-PTH-99/13, LBNL-43085, hep-ph/9904236, UC Berkeley and LBNL (1999).

[18] G. Gounaris *et al.*, in *Physics at LEP2, Vol. 1*, edited by G. Altarelli, T. Sjöstrand, and F. Zwirner (CERN, Geneva, 1996), No. CERN 96-01, p. 525.

[19] S. Weinberg, *The Quantum Theory of Fields. Volume I: Foundations* (Cambridge University Press, Cambridge — New York — Melbourne, 1995).

[20] S. Weinberg, *The Quantum Theory of Fields. Volume II: Modern Applications* (Cambridge University Press, Cambridge — New York — Melbourne, 1996).

[21] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley Publishing Company, Reading, Mass., 1995).

[22] K. G. Wilson, Phys. Rev. **B4**, 3174 (1971); K. G. Wilson, Phys. Rev. **B4**, 3184 (1971).

[23] H. Georgi, *Weak Interactions and Modern Particle Theory* (Benjamin/Cummings, Menlo Park, CA, 1984).

[24] J. F. Donoghue, E. Golowich, and B. R. Holstein, *Dynamics of the Standard Model*, Vol. 2 of *Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology* (Cambridge University Press, Cambridge, 1992).

[25] S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2239 (1969); C. G. Callan, S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2247 (1969).

[26] N. Isgur and M. Wise, Phys. Lett. **B208**, 504 (1988); N. Isgur and M. Wise, Phys. Lett. **B232**, 113 (1989); E. Eichten and B. Hill, Phys. Lett. **B234**, 511 (1990); H. Georgi, Phys. Lett. **B240**, 447 (1990); B. Grinstein, Nucl. Phys. **B339**, 253 (1990).

[27] M. Gell-Man and Y. Ne'eman, *The Eightfold Way* (Benjamin, New York, NY, 1964).

[28] H. Georgi, Phys. Lett. **B297**, 353 (1992); T. Ohl, G. Ricciardi, and E. Simmons, Nucl. Phys. **B403**, 605 (1993).

[29] C. S. Wood *et al.*, Science **275**, 1759 (1997); S. C. Bennett and C. E. Wieman, Phys. Rev. Lett. **82**, 2484 (1999).

[30] P. Sikivie, L. Susskind, M. Voloshin, and V. Zakharov, Nucl. Phys. **B173**, 189 (1980).

[31] D. A. Kosower, Phys. Rev. **D48**, 1288 (1993).

[32] L. D. Faddeev and V. N. Popov, Phys. Lett. **25B**, 29 (1967).

[33] T. Kugo and I. Ojima, Prog. Theor. Phys. Suppl. **66**, 1 (1979).

[34] C. Becchi, A. Rouet, and R. Stora, Ann. Phys. **98**, 287 (1976).

[35] J. M. Cornwall, D. N. Levin, and G. Tiktopoulos, Phys. Rev. **D10**, 11145 (1974).

[36] M. Chanowitz, M. Golden, and H. Georgi, Phys. Rev. **D36**, 1490 (1987).

[37] C. P. Burgess and D. London, Phys. Rev. **D48**, 4337 (1993).

[38] R. Haag, Phys. Rev. **112**, 669 (1958); H. J. Borchers, Nuovo Cimento **25**, 270 (1960); D. Ruelle, Helv. Phys. Acta **35**, 34 (1962).

[39] M. S. Chanowitz and M. K. Gaillard, Nucl. Phys. **B261**, 379 (1985).

[40] C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions*, Vol. 56 of *Frontiers in Physics* (Benjamin/Cummings, Menlo Park, CA, 1983).

[41] H. Georgi, Nucl. Phys. **B363**, 301 (1991).

[42] R. Kleiss and W. J. Stirling, Nucl. Phys. **B262**, 235 (1985).

[43] F. A. Berends and W. Giele, Nucl. Phys. **B294**, 700 (1987); S. Dittmaier, Phys. Rev. **D59**, 016007 (1999).

[44] R. Gastmans and T. T. Wu, *The Ubiquitous Photon: Helicity Method for QED and QCD*, Vol. 80 of *International Series of Monographs on Physics* (Clarendon, Oxford, UK, 1990).

[45] T. Ohl, Acta Phys. Polon. **B28**, 847 (1997).

[46] D. Bardin *et al.*, Nucl. Phys. B (Proc. Suppl.) **37B**, 148 (1994).

[47] E. Boos and T. Ohl, Phys. Lett. **B407**, 161 (1997).

[48] E. Boos and T. Ohl, Phys. Rev. Lett. **83**, 480 (1999).

[49] F. A. Berends and W. Giele, Nucl. Phys. **B294**, 700 (1987); M. Mangano, S. Parke, and Z. Xu, Nucl. Phys. **B298**, 653 (1988); M. Mangano and S. Parke, Phys. Rep. **200**, 301 (1991).

[50] G. 't Hooft, Nucl. Phys. **B33**, 173 (1971); G. 't Hooft, Nucl. Phys. **B35**, 167 (1971); G. 't Hooft and M. Veltman, Nucl. Phys. **B50**, 318 (1972).

[51] O. Piguet and A. Rouet, Phys. Rep. **76**, 1 (1981); O. Piguet, *Algebraic Renormalization: Perturbative Renormalization, Symmetries and Anomalies* (Springer, Berlin, 1995).

[52] N. N. Bogoliubov and O. S. Parasiuk, Acta Math. **97**, 227 (1957); K. Hepp, Comm. Math. Phys. **2**, 30 (1966); W. Zimmermann, Comm. Math. Phys. **15**, 208 (1969).

[53] J. H. Lowenstein, Comm. Math. Phys. **24**, 1 (1971); Y.-M. P. Lam, Phys. Rev. **D6**, 2145 (1972); Y.-M. P. Lam, Phys. Rev. **D7**, 2943 (1973).

[54] W. Kilian and T. Ohl, Phys. Rev. **D50**, 4649 (1994).

[55] W. Beenakker and A. Denner, Int. J. Mod. Phys. **A9**, 4837 (1994).

[56] E. Argyres *et al.*, Phys. Lett. **B358**, 339 (1995).

[57] T. Ohl, Objective Caml program `bocages`, available from the author.

[58] B. D. McKay, Congressus Numerantium **30**, 45 (1981); B. D. McKay, *nauty Users's Guide (Version 1.5)*, Technical Report No. TR-CS-90-02, Dept. Comp. Sci., Austral. Nat. Univ. (1990).

[59] H. Anlauf *et al.*, Comput. Phys. Commun. **79**, 487 (1994); H. Anlauf *et al.*, Nucl. Phys. B (Proc. Suppl.) **37B**, 247 (1994); H. Anlauf, H.-D. Dahmen, P. Manakos, and T. Ohl, *WOPPER, Version 1.5: A Monte Carlo Event Generator for* $e^+e^- \to (W^+W^-) \to 4f + n\gamma$ *at LEP2 and beyond*, Technical Report No. IKDA 96/15, hep-ph/9605457, TU Darmstadt, Darmstadt (1996).

[60] H. Phillips, private communication, 1994.

[61] N. Metropolis *et al.*, J. Chem. Phys. **21**, 1087 (1953).

[62] C. Itzykson and J.-M. Drouffe, *Statistical Field Theory*, *Cambridge Monographs on Mathematical Physics* (Cambridge University Press, Cambridge, 1989).

[63] I. Montvay and G. Münster, *Quantum Fields on a Lattice*, *Cambridge Monographs on Mathematical Physics* (Cambridge University Press, Cambridge, 1994).

[64] K. Mönig, private communication, 1999.

[65] J. Hoogland, F. James, and R. Kleiss, Comput. Phys. Commun. **99**, 180 (1996).

[66] L. Devroye, *Non-Uniform Random Variate Generation* (Springer, New York, 1986).

[67] R. Kleiss and R. Pittau, Comput. Phys. Commun. **83**, 141 (1994).

[68] F. Berends, R. Kleiss, and R. Pittau, Comput. Phys. Commun. **85**, 437 (1995).

[69] A. Denner, S. Dittmaier, M. Roth, and D. Wackeroth, *Predictions for All Processes* $e^+e^{\to}4f + \gamma$, Technical Report No. BI-TP-99-10, hep-ph/9904472 (1999).

[70] M. Roth and D. Wackeroth, private communication, 1999.

[71] G. P. Lepage, J. Comp. Phys. **27**, 192 (1978); G. P. Lepage, Technical Report No. CLNS-80/447, Cornell (1980).

[72] G. I. Manankova, A. F. Tatarchenko, and F. V. Tkachov, *MILXy Way: How Much Better than VEGAS Can One Integrate in Many Dimensions?*, Technical Report No. FERMILAB-Conf-95/213-T, FNAL, contribution to AINHEP-95, Pisa, Italy, April 1995 (1995).

[73] M. Bern and D. Eppstein, in *Computing and Euclidean Geometry*, edited by D.-Z. Du and F. Hwang (World Scientific, Singapore – New Jersey – London – Hong Kong, 1992), p. 23.

[74] D. E. Knuth, *Axioms and Hulls*, Vol. 606 of *Lecture Notes in Computer Science* (Springer, Berlin – Heidelberg – New York, 1992).

[75] T. Ohl, Fortran program `TRAMP`, unpublished.

[76] J. Ruppert, Journal of Algorithms **18**, 548 (1995).

[77] E. Byckling and K. Kajantie, *Particle Kinematics* (John Wiley & Sons, London, New York, Sydney, Toronto, 1973), p. 209.

[78] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipies in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992); W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipies in Fortran77: The Art of Scientific Computing*, Vol. 1 of *Fortran Numerical Recipies*, 2nd ed. (Cambridge University Press, Cambridge, 1992); W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipies in Fortran90: The Art of Parallel Scientific Computing*, Vol. 3 of *Fortran Numerical Recipies* (Cambridge University Press, Cambridge, 1996).

[79] T. Ohl, Fortran program `VAMP`, available from the author.

[80] T. Ohl, Comput. Phys. Commun. (1999), in print, hep-ph/9806432.

[81] R. Krecker, Comput. Phys. Commun. **106**, 258 (1997).

[82] S. Veseli, Comput. Phys. Commun. **108**, 9 (1998).

[83] D. E. Knuth, *Seminumerical Algorithms*, Vol. 2 of *The Art of Computer Programming*, 3rd ed. (Addison-Wesley, Reading, MA, 1998).

[84] R. Kleiss, W. J. Stirling, and S. D. Ellis, Comput. Phys. Commun. **40**, 359 (1986).

[85] R. Kleiss and W. J. Stirling, Nucl. Phys. **B385**, 413 (1992).

[86] K. Hagiwara, R. D. Peccei, D. Zeppenfeld, and K. Hikasa, Nucl. Phys. **B282**, 253 (1987).

[87] L. Durand, P. DeCelles, and R. Marr, Phys. Rev. **126**, 1882 (1962); K. J. Kim and Y.-S. Tsai, Phys. Rev. **D7**, 3710 (1973).

[88] W. Buchmüller and D. Wyler, Nucl. Phys. **B268**, 621 (1986).

[89] A. Nippe and T. Ohl, Study for the TESLA conceptual design report, unpublished.

[90] J. Erler and P. Langacker, Phys. Rev. **D52**, 441 (1995).

[91] A. Falk, M. Luke, and E. H. Simmons, Nucl. Phys. **B365**, 523 (1991).

[92] B. Holdom, Phys. Lett. **B258**, 156 (1991).

[93] T. Appelquist and G.-H. Wu, Phys. Rev. **D48**, 3235 (1993).

[94] A. Nippe, private communication, 1996; C. Burgard and S. Petzold, private communication, 1999.

[95] E. Accomando and A. Ballestrero, Comput. Phys. Commun. **99**, 270 (1997).

[96] W. Beenakker *et al.*, in *Physics at LEP2, Vol. 1*, edited by G. Altarelli, T. Sjöstrand, and F. Zwirner (CERN, Geneva, 1996), No. CERN 96-01, p. 79.

[97] D. Bardin *et al.*, in *Physics at LEP2, Vol. 2*, edited by G. Altarelli, T. Sjöstrand, and F. Zwirner (CERN, Geneva, 1996), No. CERN 96-01, p. 3.

[98] A. Ballestrero, private communication, 1999.

[99] E. Boos, in *Physics and Exerimentation with Future Linear e^+e^--Colliders, Sitges, Spain* (World Scientific, Singapore – New Jersey – London – Hong Kong, 1999), (to be published).

[100] P. Chen and K. Yokoya, Phys. Rev. Lett. **61**, 1101 (1988); P. Chen, Phys. Rev. **D46**, 1186 (1992).

[101] P. Chen *et al.*, Nucl. Inst. Meth. **A355**, 107 (1995).

[102] D. Schulte, Simulation code Guinea Pig, 1996, available from the author.

[103] T. Ohl, Comput. Phys. Commun. **101**, 269 (1997).

[104] T. Ohl, Fortran library `circe`, updated versions available from the author.

[105] E. Boos *et al.*, Phys. Rev. **D57**, 1553 (1998).

[106] E. Accomando, A. Ballestrero, and M. Pizzio, Nucl. Phys. **B512**, 19 (1998).

[107] F. Gangemi *et al.*, Eur. Phys. J. **C9**, 31 (1999); F. Gangemi *et al.*, *Top Quark Physics in Six Quark Final States at the Next Linear Collider*, Technical Report No. FNT-T-99-08, hep-ph/9905271, Ferrara University.

[108] G. 't Hooft and M. Veltman, Nucl. Phys. **B44**, 189 (1972).

[109] D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973); H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).

[110] S. Weinberg, Physica **96A**, 327 (1979).

[111] J. Polchinski, Nucl. Phys. **B231**, 269 (1984).

[112] J. Polchinski, *Effective Field Theory and the Fermi Surface*, Technical Report No. NSF-ITP-92-132, UTTG-20-92 (1992).

[113] S. Coleman and E. Weinberg, Phys. Rev. **D7**, 1888 (1973).

[114] A. Cabibbo, L. Maiani, G. Parisi, and R. Petronzio, Nucl. Phys. **B158**, 295 (1979).

[115] B. J. Warr, Ann. Phys. **183**, 1 (1988); B. J. Warr, Ann. Phys. **183**, 59 (1988).

[116] S. Weinberg, Phys. Rev. **D8**, 3497 (1973).

[117] V. V. Sudakov, Sov. Phys.—JETP **3**, 65 (1956).

[118] H. Georgi, Annu. Rev. Nucl. Part. Sci. **43**, 209 (1993).

[119] T. Appelquist and J. Carazzone, Phys. Rev. **D11**, 2856 (1975).

[120] T. Inami and C. S. Lim, Prog. Theor. Phys. **65**, 297 (1981).

[121] A. Manohar and H. Georgi, Nucl. Phys. **B234**, 189 (1984); H. Georgi and L. Randall, Nucl. Phys. **B276**, 241 (1986); H. Georgi, Phys. Lett. **B298**, 187 (1993).