Unification: the fabric of understanding Nature

Just as done several times in the previous chapters, we reconsider the historical development of the key aspects of modern physics, using the benefit of hindsight to perceive the character of this development. Throughout the foregoing material, the Democritean atomism provided the *warp*, complemented by the gauge principle as its *weft*. Along the way, however, this fabric reveals the ubiquity of the third conceptual strand (*woof*, as it were [☞ lexicon entry on p. 508, in Appendix B.1]) – unification; we now turn to explore this more closely.

8.1 Indications

The Newtonian theory of gravity unites the mechanics of the so-called terrestrial and the celestial objects and so eradicates this difference postulated within the Aristotelian philosophy of Nature, which the Roman Catholic clergy (sanctioned by the AD 313 Edict of Milan) imposed as exclusive of all other world views from the pre-hellenic and the hellenic cultures. According to Newton's law of gravity – as a descriptive model of this natural phenomenon – gravity is obeyed equally by both the Sun and the Moon, and the planets and the stars, by the communication satellites and the rockets as well as the Rockettes, by both the basketball and the baseball balls as well as the players in those games, and of course also by the apple that supposedly fell from the tree under which Newton sat...

Of course, Newton's unified description of Nature is not a unification of pre-existing theoretical models – in the contemporary sense of the word "model." It is, however, one of the first rigorous applications of the principle that Nature is one and that it can be understood in a unified fashion, and not as a (jury rigged) patchwork of different and diverse ideas – each with but a very narrow aim and applicability.

It behooves us then to examine also the nature of our ideas about unification.

8.1.1 Unification of relativistic and quantum physics

Modern fundamental physics is based on the requirement that a description of Nature include both its quantumness and its relativity, in the senses of the general theory of relativity.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

Special relativistic unification

The first example of unification of *existing* scientific models is provided by the Maxwell equations: Indeed, Ampère's and Faraday's laws and Gauss's laws (5.72) were already known, as well as experimentally verified in situations for which those laws of Nature had been identified.[1] Their combination into one unified, *electromagnetic system* – and the extension of Ampère's law for the sake of agreement with the continuity equation and general consistency in the non-static/stationary case – has far-reaching consequences:

1. The electric field and the magnetic field (viewed as two distinct physical phenomena) are the limiting cases of a unified electromagnetic field, in the formal limit $c \to \infty$; accordingly, the Maxwell equations (5.72), without magnetic (monopole) charges and currents, become (using the relation $1/\epsilon_0 c^2 = \mu_0$)

$$\vec{\nabla}\cdot\vec{E} = \frac{4\pi\rho_e}{4\pi\epsilon_0}, \qquad \vec{\nabla}\times\vec{B} = \frac{4\pi}{4\pi\epsilon_0 c^2}\vec{j}_e + \frac{1}{c^2}\frac{\partial\vec{E}}{\partial t} \;\to\; \mu_0\vec{j}_e, \tag{8.1a}$$

$$\vec{\nabla}\cdot\vec{B} = 0, \qquad -\vec{\nabla}\times\vec{E} = \frac{\partial\vec{B}}{\partial t}. \tag{8.1b}$$

   Thus, in the formal limit $c \to \infty$, only the last relation (Faraday's law) still relates the electric and the magnetic fields, and only when the magnetic field varies in time and the electric field varies in space, so as to have a nonzero curl: $\vec{\nabla}\times\vec{E} \neq 0$. In turn, the full electrodynamics (5.72) is then the extension of this electro-and-magneto-*static* system, both self-consistent and consistent with Nature.
2. Changes in the electromagnetic field propagate with the speed of light, in the form of waves. Using the Lorenz gauge, the Maxwell equations (5.72) without magnetic (monopole) charges and currents produce the wave equation for the 4-vector gauge potential (5.92), so that their changes propagate at the speed of light in vacuum, c. The electromagnetic field, as gauge-invariant derivatives of the gauge potentials (5.15), also satisfies the wave equation

$$\Box\,\vec{B} = \vec{\nabla}\times(\mu_0\vec{j}_e), \qquad \Box\,\vec{E} = -\vec{\nabla}\left(\frac{\rho_e}{\epsilon_0}\right) - \frac{\partial(\mu_0\vec{j}_e)}{\partial t}. \tag{8.2}$$

3. The system of Maxwell equations (5.72) has symmetries:
   (a) Lorentz transformations of spacetime (3.1), i.e., (3.13) and corresponding transformations of the electromagnetic field (5.75),
   (b) duality (5.86) between the electric and the magnetic field.
4. The existence of magnetic (monopole) charges and currents would obstruct the (unambiguous) expression of the electromagnetic field in terms of a gauge 4-vector potential [☞ Comment 5.6 on p. 185].
5. The regime where the unified electrodynamics may be regarded as a collection of separate subjects of electro-*statics*, magneto-*statics* and wave optics is the "$c \to \infty$" formal limit.

Comment 8.1 *Since c is a natural constant, the formal limit "$c \to \infty$" makes sense only in the form of dimensionless ratios $v_{ij}/c \to 0$, where v_{ij} ranges over all relative speeds observable in the considered system. This has three significant consequences:*

1. *Non-relativistic physics is a special,* ***limiting case*** *of relativistic physics, which is in turn an* ***extension*** *of non-relativistic physics. For any given system, in the*

[1] Maxwell noticed that without the displacement current, $-\mu_0\partial(\epsilon_0\vec{E})/\partial t$, the divergence of Ampère's original law, $\vec{\nabla}\times\vec{B} = \mu_0\vec{j}_e$, produces $\vec{\nabla}\cdot\vec{j}_e = 0$, which holds only in the restricted cases when the free charge density in the entire observed space is unchanging in time, i.e., only in the static/stationary situations for which Ampère originally identified the law.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

space of all possible relative speeds $\{v_{12}, v_{13}, \dots\}$*, the non-relativistic regime involves only the lowest-order nonzero results in the* $v_{ij}/c \ll 1$ *approximation, i.e.,* ***near the point****:* $v_{ij} = 0$ *for all* i, j*; everything else is relativistic physics.*

2. *By non-relativistic systems one may understand only the cases where the relativistic corrections are* ***negligible*** *– for which the limits of precision are necessarily subject to* ***convention****.*
3. *Since the changes in the electromagnetic field propagate at the speed of light, all systems with variable electromagnetic fields are unavoidably relativistic.*

The property that changes in the electromagnetic field propagate as waves, at the speed of light, unifies the (electro- and magneto-)*static* phenomena with the wave phenomena (ultraviolet radiation, light, heat radiation and radio-waves, which were known by the end of the nineteenth century to be but different types of electromagnetic radiation), and then also the high-frequency limit of wave optics known as geometric optics.

Digression 8.1 From the contemporary, symmetry vantage point, the symmetries of the Maxwell equations are the Lorentz transformations [☞ Section 3.1]. The symmetries of Newtonian mechanics are the Galilean transformations, which differ from Lorentz transformations in that the boost transformations do not change time:

$$\textbf{Galileo} \quad \vec{r}\,' = \vec{r} - \vec{v}t, \qquad t' = t, \tag{8.3a}$$

$$\textbf{Lorentz} \quad \vec{r}\,' = \vec{r} - \gamma\vec{v}t + (\gamma - 1)(\hat{v}\cdot\vec{r})\,\hat{v}, \qquad t' = \gamma\Big(t - \frac{\vec{v}\cdot\vec{r}}{c^2}\Big). \tag{8.3b}$$

In Newtonian physics, time is absolute. Since charged particles interact with the electromagnetic fields and when they move, it is necessary that the theoretical model of those interactions is a single, coherent and consistent theoretical system – which can happen only if one can either:

1. adapt the Maxwell equations so as to exhibit Galilean symmetries of Newtonian physics,
2. or adapt Newtonian laws so as to exhibit Lorentz symmetries of relativistic physics.

As is well known, Nature picks the second, and not the first of these logical possibilities.

General relativistic unification

Chapter 9 will provide a telegraphic review of the general theory of relativity, but let us note here that the "general theory of relativity" (and then also its special case, the special theory of relativity) is in fact a ***theoretical system*** [☞ Section 8.3.1]. The pivotal idea in the theory of relativity is also the gauge principle, but applied to the "real," i.e., concrete spacetime, rather than to an abstract space of phases as was the case with electroweak and strong interactions [☞ Chapter 5]:

1. To describe physical systems, one uses coordinate systems the points of which are the points of spacetime in which the parts of that system move. To this end, one uses the 4-vector of spacetime coordinates, x.
2. The coordinates in such coordinate systems are not themselves physically observable, i.e., they cannot be measured. Indeed, absolute positions of various objects cannot be measured, but distances between them can.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

3. To measure distances,

$$s(\mathrm{x}_i, \mathrm{x}_f) := \int_{\mathrm{x}_i}^{\mathrm{x}_f} \mathrm{d}s, \qquad \text{where} \quad \mathrm{d}s^2 := g_{\mu\nu}(\mathrm{x})\, \mathrm{d}x^\mu \mathrm{d}x^\nu, \tag{8.4}$$

one must know the metric tensor $g(\mathrm{x})$ in the chosen coordinate system, represented by $\mathrm{x} = (x^0, x^1, x^2, x^3)$. The components $g_{\mu\nu}(\mathrm{x})$ are – in principle, and definitely in the general case corresponding to the *general* theory of relativity – arbitrary functions of the coordinates x. For special relativity, $g_{\mu\nu}(\mathrm{x}) = -\eta_{\mu\nu}$; see (3.17)–(3.19).

4. Since the coordinates x cannot be observed directly, it ought be possible to change the coordinate system – through the substitution $\mathrm{x} \to \mathrm{y}$, but so that

$$\mathrm{d}s^2_{(x)} = g_{\mu\nu}(\mathrm{x})\, \mathrm{d}x^\mu \mathrm{d}x^\nu \overset{!}{=} g_{\mu\nu}(\mathrm{y})\, \mathrm{d}y^\mu \mathrm{d}y^\nu = \mathrm{d}s^2_{(y)}, \tag{8.5}$$

from which it follows that [☞ Digression 3.2 on p. 88, and Chapter 9]

$$g_{\mu\nu}(\mathrm{y}) = \frac{\partial x^\rho}{\partial y^\mu} \frac{\partial x^\sigma}{\partial y^\nu}\, g_{\rho\sigma}(\mathrm{x}), \tag{8.6}$$

that is, that the metric tensor is indeed a tensor, of rank 2 and of type $(0, 2)$.

5. Chapter 9 shows how invariance with respect to general coordinate transformations implies the existence of gauge potentials and the gravitational interaction – exactly the way invariance with respect to local phase transformations implies the Yang–Mills type gauge interactions [☞ Chapters 5–6].

Comment 8.2 *For the special theory of relativity, we have $g_{\mu\nu}(\mathrm{x}) \to -\eta_{\mu\nu}$,[2] which is the constant metric tensor (3.19) of the "flat" spacetime. In this sense, the special theory of relativity is a "**limit-point**" in the space of all possible general coordinate systems and corresponding metric tensors described in the general theory of relativity. In turn, the general theory of relativity is then an **extension** of the special theory of relativity.*

The practical demarcation between special and general theories of relativity may thus naively be estimated by considering the departure of the actual metric tensor $g_{\mu\nu}(\mathrm{x})$ from the metric tensor of flat spacetime, $-\eta_{\mu\nu}$. This, however, is not well defined. Indeed, owing to the relation (8.6), neither is specifying any particular component of the metric tensor nor is its comparison with the same component from another metric tensor independent of the choice of coordinates. However, there do exist so-called curvature invariants, the values of which are independent of coordinate choices, and these then may serve for demarcation purposes. In $(3+1)$-dimensional spacetime, there are 20 such invariants, of which the simplest one is the so-called scalar curvature, $R := g^{\mu\rho} R_{\mu\nu\rho}{}^{\nu}$, where $R_{\mu\nu\rho}{}^{\sigma}$ is the so-called Riemann tensor [☞ Chapter 9]. Suffice it to say, if any one of these 20 curvature invariants cannot be neglected (in the considered processes and in comparison with some earlier specified precision limits), the system is generally relativistic.

The general theory of relativity contains (Einstein's) model of gravity, while the special theory of relativity pertains to flat spacetime, with no gravitational effects. Thereby, the special theory of relativity may be regarded as the formal $G_N \to 0$ limit of the general theory.

Comment 8.3 *As in the case of the formal limit "$c \to \infty$" and since G_N is a natural constant, the formal limit "$G_N \to 0$" may be understood only as a statement that all characteristic quantities of the system commensurate with G_N (of the same physical units) are much larger*

[2] The expression (8.4) defines the metric tensor $g_{\mu\nu}$ by way of defining the **distance**, while the expression (3.17) defines the proper **time** in spacetime. The signs in $\eta_{\mu\nu}$ are therefore opposite from the signs in $g_{\mu\nu}$.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

than G_N. Intuitively, these characteristic quantities ought to be some invariant measures of the spacetime curvature, but all such invariants are computable from the Riemann tensor, the dimensions of which are $[R_{\mu\nu\rho}{}^{\sigma}] = L^{-2}$. On the other hand, $[G_N] = L^3T^{-2}M^{-1}$ and curvature invariants (obtained as various contractions of various tensor products of the Riemann tensor) cannot be compared with G_N, but can be compared with the constant $\ell_P := \sqrt{\hbar G_N/c^3}$, the Planck length [☞ Table 1.1 on p. 24].

Thus, for the purposes of estimating the "non-gravitational" limiting case, it is more convenient to use the natural constant ℓ_P ***instead of*** *Newton's gravitational constant. This limiting case then may be written formally as "$\ell_P \to 0$," understanding here relations of the type $|\boldsymbol{R}_i|\ell_P \ll 1$, where:*

1. *$|\boldsymbol{R}_i|$ is the norm of the ith curvature invariant, defined so as to have dimensions $[\boldsymbol{R}_i] = L^{-1}$;*
2. *the relation "$\ll$" here means "smaller than a previously set limit of precision."*

In this sense, the notation "$G_N \to 0$" is being used as a synonym for the formal limit "$\ell_P \to 0$", while keeping $\hbar$ and c constant [☞ Comments 8.1 on p. 294 and 8.4 on p. 298].

Quantum unification

As the Maxwell equations – the theoretical model of the electromagnetic field – indicate that the electro-static and the magneto-static fields are only limiting cases of the electromagnetic field whereby the descriptions of these natural phenomena are unified, so does quantum mechanics unite the notion of a particle and that of a wave.

The very notion of a particle presupposes that the position of the observed object in "ordinary" space may be localized arbitrarily well, i.e., that the object is ideally located in a perfectly well-specified (mathematically dimensionless) point of "ordinary" space: the position of this object is perfectly precisely specified. In a complementary fashion, the very notion of a (plane) wave presupposes that the position of the observed object in *momentum space* may be localized arbitrarily well, i.e., that the object is ideally located in a perfectly well-specified (mathematically dimensionless) point of momentum space: the wave vector of the object is perfectly precisely specified.

However, the Heisenberg indeterminacy relations, $\triangle x \triangle p_x \geqslant \frac{1}{2}\hbar$, imply that a *quantum* object cannot be localized more precisely than within a region in the phase space,[3] the "surface area" of which is never smaller than $\frac{1}{2}\hbar$. This gives the phase space in quantum physics a "granular" structure. In turn, it is also known that functions (or, more generally, distributions) over the phase space that may be used to represent *classical* observables cannot reproduce consistently and completely all properties of the *quantum* state operator; see quantum mechanics textbooks such as Ref. [29]. Thus, quantum physics cannot be described simply as classical physics with the additional requirement of a "granular" phase space. Quantum mechanics teaches us that real "things" are neither ideal particles nor ideal waves, but "something else"; something that in appropriate circumstances may be *approximated* by the limiting case of a point-particle, while in other circumstances an *approximation* by the limiting case of a wave is more precise.

The conceptual analogy with electro-*static* and magneto-*static* fields on one hand, and electromagnetic waves (always moving) on the other should be manifest. It should then come as no surprise that field theory in this conceptual sense interpolates between particles and waves. However, field theory is not a theory of a collection of wave packets – that literally interpolate between particles represented by the Dirac δ-function as one limiting case, and plane waves as the other

[3] The geometric shape, and even connectedness of this region remains a-priori undetermined, regardless of the choice of a system, and its evolution during the passage of time.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

limiting case. Field theory contains both wave packets as limiting cases – less special than particles and plane waves, but limiting cases nevertheless.[4]

Finally, the transition from quantum to classical physics is often cited as the formal limiting process $\hbar \to 0$, which identifies classical physics as a limiting case of quantum physics.

> ***Comment 8.4*** *Since $\hbar$ is a natural constant, this limiting process makes sense only as the limit $(\hbar / S_i) \to 0$ for $i = 1, 2 \ldots$, where S_i are various physical observable quantities characteristic for the given system and with units $\frac{ML^2}{T}$, such as the angular momentum and Hamilton's action, $S = \int dt\, L$, where L is the Lagrangian of the system. In this precise sense, classical (non-quantum) physics is a* ***limiting case*** *of quantum physics, which is in turn an* ***extension*** *of classical physics.*

The theoretical system of relativistic quantum physics

The combination of limiting processes described in Comments 8.1, 8.2 and 8.4 in this section then provides the complete depiction (see Figure 8.1) of the *theoretical system* within which the Standard Model of elementary particle physics is formulated [☞ Section 8.3].

For all Standard Model purposes, suppose that the spacetime curvature and corresponding gravitational effects are negligible, i.e., that a full sequence of conditions of the form $\boldsymbol{R}_i \ell_P \to 0$ is satisfied, as discussed in Comment 8.3 [☞ also Chapter 9, as well as Refs. [508, 62, 367, 548, 66]], and that reduces Einstein's general theory of relativity to the special theory of relativity with no gravitation; the Newtonian theory of gravity may be derived as a lowest-order nonzero effect *near* this limit [95, 96, 271, 58]; see also Section 9.2.4. In individual interaction processes between elementary particles, the gravitational interaction is many orders of magnitude weaker than the strong or even the electroweak interactions, whereby special relativity suffices for all Standard Model purposes.

With this assumption, the schematic diagram in Figure 8.1 reduces to the first quadrant in the coordinate $(\frac{1}{c}, \hbar)$-plane, which represents (specially) relativistic quantum physics, i.e., field theory.

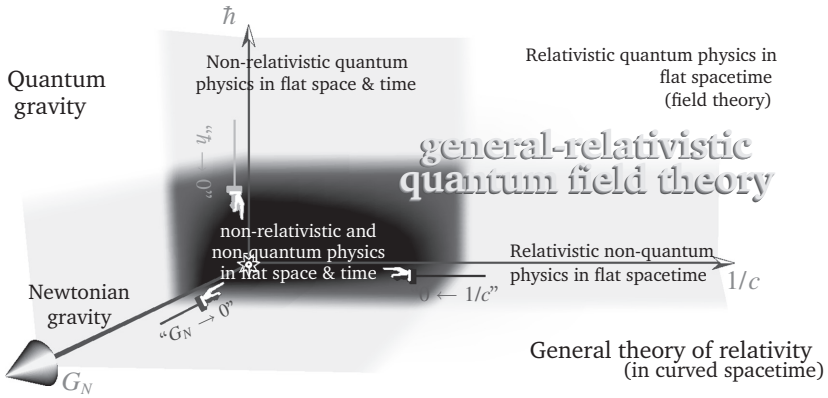

Figure 8.1 A sketch of the limiting cases of the general and special theory of relativity as well as quantum physics. The boundaries of the formal transitions into the approximations "$c \to \infty$," "$\hbar \to 0$" and "$G_N \to 0$" (i.e., "$\ell_P \to 0$"; see text) are conventionally defined, as depicted by gradual shading. General-relativistic quantum field theory is up front, well inside the first octant.

[4] It may help to imagine the palette of possibilities covered by field theory as a multi-dimensional geometric object with an "edge." The points of this "edge" correspond to various wave-packets, and its two end-points correspond to the particle and the plane wave, respectively.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

The demarcations that determine the negligibility of characteristic quantities $(\hbar/S_i)$ and (v_{jk}/c) of the system are *conventional*, and this is represented in Figure 8.1 as a gradual change in the shading. Here, the non-relativistic physics is "sufficiently near" the vertical axis, and non-quantum physics is "sufficiently near" the horizontal axis. The practical criteria for this "nearness" – i.e., the boundary where the non-relativistic or non-quantum approximation is no longer sufficiently good – depends on the adopted *conventions* regarding the required precision of computational results.

Let us then emphasize the conceptual differences:

1. in phase transitions, the boundary between the symmetric and the non-symmetric phase is precisely determined by the system: see Conclusion 7.5 and relation (7.45);
2. the transition from quantum (or relativistic) physics into the non-quantum (non-relativistic) *approximation* is conditioned by the convention of computation precision. Strictly speaking (with absolute precision), non-quantum and non-relativistic physics are merely idealized limiting cases.

The transition from the regime where the electroweak interaction is united into the regime where electrodynamics essentially differs from weak interactions (photons are massless, $W^{\pm}$-, Z^0-bosons are massive) is manifestly a phase transition and not a conventional approximation. In turn, the transition from the regime of electrodynamics into the regime where we – practically and pragmatically – separate electro-statics from magneto-statics is conditioned by the convention of computational precision, i.e., whether or not relativistic corrections may be neglected.

However, there do exist significant similarities. The conceptual similarity is reflected in the facts that both electrodynamics and electroweak interactions have both a "unified" and a "separated" regime, as well as that the symmetries of the system in the unified regime are larger than the symmetries in the separated regime; see Table 8.1.

Table 8.1 Conceptual similarities and differences between the unification of the electric and the magnetic fields into the electromagnetic (EM) one, and the electromagnetic and weak fields into the electroweak (EW) field. $Po(1,3)$ is the Poincaré group of linear transformations of spacetime: Lorentz transformations and translations.

	United regime	**Separated regime**
Electromagnetism	The relative speed between at least two subsystems is not negligibly small, $v_{ij}/c \not\ll 1$.	The relative speed between at least two subsystems is negligibly small, $v_{ij}/c \ll 1$.
	The transition demarcation is ***specified by a convention*** in resolution.	
	Separation and differentiation between the $\vec{E}$- and the $\vec{B}$-fields depends on the choice of the coordinate system; see Example 5.1 on p. 183, and relations (5.75) and (5.77).	In a system where the free charges are static and the idealized currents stationary, the electric and the magnetic fields are static and perfectly separated.
	The symmetries of the Maxwell equations form the Lorentz group, together with spacetime translations, i.e., the Poincaré group, $Po(1,3)$.	The symmetries of electro- and magneto-static systems are limited to rotations in space, Galilean boosts and translations in space and time, $Ga(1,3) \subsetneq Po(1,3)$.
Electroweak int.	Particles in a process have energies $E_i > \hbar c\sqrt{\lambda}\langle \mathbb{H}\rangle|_{\varkappa<0} \sim M_{W^{\pm}}c^2$.	Particles in a process have energies $E_i < \hbar c\sqrt{\lambda}\langle \mathbb{H}\rangle|_{\varkappa<0} \sim M_{W^{\pm}}c^2$.
	The transition demarcation (the order parameter critical value) is ***determined by the system***.	
	$W^{\pm}, W^3_\mu$ and B_μ are the normal modes, and are all massless.	B_μ and W^3_μ are not normal modes; A_μ (massless) and Z_μ (massive) are; see relations (7.85)–(7.86).
	Local (gauge) symmetries of electroweak interactions form the $SU(2)_w \times U(1)_y$ group.	Local (gauge) symmetries of electroweak interactions reduce $U(1)_Q \subset SU(2)_w \times U(1)_y$.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

Conclusion 8.1 (unification) *Since Newton's* Principia *(1687) and through the unification of electroweak interactions (Glashow, Weinberg and Salam, 1979 Nobel Prize), three distinct notions of unification have grown into fundamental physics:*

- **(a) conceptual** *in the sense that Nature is one and that its scientific descriptions (models) should be conceptually uniform, and not a patchwork (hodgepodge) of diverse and disparate ideas;*
- **(b) limiting** *in the sense that one marked "regime" of behavior of a system is, strictly speaking, merely a special limiting case (i.e., approximation) of another, more general and/or more exact description;*
- **(c) phase/regime** *where the description of a system contains a definition of an order parameter and its critical value that divides two phases, i.e., regimes of a system.*

Note the double duty pulled by the word "regime," used in two different senses in the second and third notions of unification as listed here. Similarly, the word "phase" is used here in the sense exemplified by solids vs. liquids – very different from its use in Chapters 5–7.

8.1.2 Indications for exploring beyond the Standard Model

The Standard Model explains a lot, but also indicates the unknown source of some of the basic characteristics of this model and the state of understanding Nature that this model represents:

Spacetime For Standard Model purposes, one *assumes* the spacetime to be a continuous topological real 4-dimensional space with a flat metric tensor $-\eta_{\mu\nu}$ of signature $(1,3)$, i.e., that one of the four dimensions is of a time-like and three are of a space-like character. We do not know why this is so.

The interaction hierarchy The fundamental interactions in the Standard Model emerge from the gauge principle and the local (gauge) symmetry group $SU(3)_c \times SU(2)_w \times U(1)_y$. The dependence of the interaction strength on the 4-momentum transfer involved where this strength is measured as well as the electroweak symmetry-breaking $SU(2)_w \times U(1)_y \to U(1)_Q$ are described within the Standard Model. However, the relative intensities of the concrete values of the parameters α_s, α_w and α_y (i.e., α_e) – obtained by measuring at any one concrete energy – are not determined within the Standard Model and may only be regarded as given (and unexplained) "initial data."

The scale and the mass hierarchy structure All Standard Model fermions acquire their mass via interaction with the Higgs field, through the field shift $\mathbb{H} \to \mathbb{H} + \langle \mathbb{H} \rangle$ [☞ relations (7.109)–(7.113)]. However, nothing in the Standard Model determines the concrete values of the specific constants h_Ψ that describe the intensity of the direct (Yukawa) interaction of the Standard Model fermions with the Higgs boson – and thus also the masses of these fermions [☞ Tables 4.1 on p. 152 and C.2 on p. 526]. Since $\langle \mathbb{H} \rangle$ is determined from the experimental data for $m_Z = 91.187\,6\,\mathrm{GeV}/c^2$ [☞ relation (7.81) and (7.86)], it follows that

$$\langle \mathbb{H} \rangle \sim 10^2\,\mathrm{GeV}/c^2, \quad \text{and} \quad h_u, h_d \sim 10^{-5}, \quad h_s \sim 10^{-3}, \quad h_c, h_b \sim 10^{-2}, \quad h_t \sim 1. \tag{8.7}$$

Neither the general smallness h_Ψ (except h_t) nor the hierarchy of these parameters is explained in the Standard Model. Until the Higgs particle is fully confirmed and its characteristics (including all the coupling parameters h_Ψ) are measured, the fermion masses remain without explanation in the Standard Model.

CKM quark mixing The very fact that the eigenstates of the free Hamiltonian are also the eigenstates of the strong, electromagnetic and gravitational interactions, but not also of weak interaction is not unusual: there is no a-priori theoretical reason for a coincidence of eigenstates of all various

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

interaction terms in the Hamiltonian. However, the origin and the concrete values of the Cabibbo–Kobayashi–Maskawa parameters (angles) that control the quark mixing in weak interactions (2.53) are not determined at all by the Standard Model and remain unknown☝.

Neutrino mixing and oscillations Similarly to quarks, there is no a-priori theoretical reason for a coincidence of the eigenstates of the free and the (only) weak interactive term in the Hamiltonian for neutrinos. However, the origin and the concrete value of the parameter M_ν in equation (7.132) and, more generally, the origin and the concrete values of the parameters (in the PMNS-matrix [☞ Section 7.3.2]) that control the neutrino mixing in free propagation (as compared the neutrinos defined by the weak interactions) are also not determined at all by the Standard Model, and this remains an open problem☝.

The number of fermion families The Standard Model simply includes the fact that there exist three families of fundamental fermions [☞ Table 7.1 on p. 275], but this fact is neither mandated nor explained and remains one of the puzzles of the Standard Model☝.

***CP*-violation** The combined discrete *CPT*-operation must be a symmetry in all Lorentz-invariant models [☞ Section 4.2.3]. However, the combined *CP*-operation need not be (and is not) a symmetry of Nature, and neither need then the time reversal operation be. On the other hand, *T*-violation is necessary for the irreversible creation of a sufficient surplus of matter (as compared to antimatter) in the first seconds of the Big Bang, and *CP*-violation via weak interactions is, roughly and little as it is, of just the sufficient amount. However, nothing in the Standard Model explains the concrete value of the angle δ_{13} in the CKM matrix (2.53), nor the complete absence of the otherwise perfectly possible – and many orders of magnitude larger – *CP*-violation through strong nuclear interactions [☞ Section 6.3.1], which remains a complete mystery☝.

Cosmological constant Phase transitions always have excess energy density [☞ Conclusion 7.1 on p. 258 and Comment 7.3 on p. 265]. For water to freeze, an external heat reservoir must remove this excess energy. However, when the entire Universe undergoes a phase transition, there is no "external heat reservoir," and this energy remains as a homogeneous and isotropic background energy. The recent discovery that the expansion of the Universe is in fact accelerating implies the existence of some kind of background "*dark energy*" – however, the observed value of even the so-called cosmological constant is many tens of orders of magnitude smaller than the excess energy density of the electroweak phase transition; the origin and the concrete value of this astoundingly extravagant discrepancy remains a puzzle; see Comment 7.3 on p. 265☝.

Dark matter Observations of the distribution of rotation speeds of stars about their galactic centers imply the existence of an invisible source of gravity (mass), the quantity and volume of which surpasses the mass and volume of the visible matter in galaxies. The Standard Model contains no adequate candidate for such matter, the origin and nature of which then remain a puzzle☝. The variants of the cosmological "inflationary model" require that the total amount of matter in the Universe should even be ten times more than the best estimates for the amount of visible matter. For these models – which successfully describe most cosmically large-scale properties of our Universe – the existence of dark matter is crucial.

— ❦ —

The questions that the Standard Model uncovers may in many cases be formulated only based on the description of Nature and insights into its properties given precisely by that same Standard Model. It is then not inappropriate to regard the Standard Model as a tool for systematizing our questions about Nature that are conceptually beyond reach of the Standard Model. We thus speak of research "beyond the Standard Model."

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

Theoretical, experimental – and even aesthetic – successes of the electro-weak unification inspired many a researcher in the last quarter of the twentieth century to formulate a model that would unify the strong with the electroweak interaction, as well as explain at least some of the Standard Model puzzles. This idea receives significant support from the fact that the coupling parameters α_s, α_w and α_y change with the magnitude of the 4-momentum transfer at which these parameters are measured – and in a, roughly, convergent fashion. That is, if we suppose that above the energies $\sim 10^2$ GeV there exist no new fundamental fermions as well as no new interactions – which is referred to as the "grand desert hypothesis" – the functions $\alpha_s(\mathrm{q})$, $\alpha_w(\mathrm{q})$ and $\alpha_y(\mathrm{q})$ converge and meet approximately at the energy $|\mathrm{q}|c \sim 10^{15-17}$ GeV. The details of this convergence and of this "merging" depend on the concrete model and additional assumptions and so are necessarily of a speculative nature; finally, one talks about an extrapolation over 15–16 orders of magnitude, with no precedent in the history of physics!

Consider the relation (5.202), as well as (6.79), which holds for the general case of $SU(n)$-gauge interactions of n_f fermion flavors, and note that the reciprocals of the fine structure parameters are approximately linear functions of the logarithm of the magnitude of the 4-momentum transfer $|\mathrm{q}|$ at which the parameters are measured:

$$\left.\begin{aligned} U(1): \quad & \frac{1}{\alpha_{1,R}(|\mathrm{q}^2|)} \approx \frac{1}{\alpha_{1,R}(\mu^2c^2)} - \frac{4}{12\pi}\ln\Big(\frac{|\mathrm{q}^2|}{\mu^2c^2}\Big) \\ SU(n): \quad & \frac{1}{\alpha_{n,R}(|\mathrm{q}^2|)} \approx \frac{1}{\alpha_{n,R}(\mu^2c^2)} + \frac{11n-2n_f}{12\pi}\ln\Big(\frac{|\mathrm{q}^2|}{\mu^2c^2}\Big) \end{aligned}\right\} \quad |\mathrm{q}^2| \gg \mu^2c^2, \tag{8.8}$$

where μ is the largest fermion mass that can occur in the loops such as (5.201), the total number of which equals n_f. At energies over $\mu c^2 = m_\tau c^2 = 174.2$ GeV, we have

$$SU(3)_c: \quad n_f = 3\times 2_{(w)}, \qquad 11n - 2\,n_f = +21, \tag{8.9}$$

$$SU(2)_w: \quad n_f = 3\times (3_{(c)}+1), \qquad 11n - 2\,n_f = -2, \tag{8.10}$$

for $SU(2)_w$ and the same μ, and where the number of $SU(3)_c$-interacting quarks equals 6 (one doublet of quark $SU(3)_c$-triplets in each of the three families) and the number of $SU(2)_w$-interacting fermions equals 12: one (color) triplet of quark $SU(2)_w$-doublets and one lepton $SU(2)_w$-doublet in each of the three families. One thus obtains

$$U(1)_y: \quad \frac{1}{\alpha_{y,R}(|\mathrm{q}^2|)} \approx \frac{1}{\alpha_{y,R}(\mu^2c^2)} - \frac{4}{12\pi}\ln\Big(\frac{|\mathrm{q}^2|}{\mu^2c^2}\Big), \tag{8.11a}$$

$$SU(2)_w: \quad \frac{1}{\alpha_{w,R}(|\mathrm{q}^2|)} \approx \frac{1}{\alpha_{w,R}(\mu^2c^2)} - \frac{2}{12\pi}\ln\Big(\frac{|\mathrm{q}^2|}{\mu^2c^2}\Big), \tag{8.11b}$$

$$SU(3)_c: \quad \frac{1}{\alpha_{s,R}(|\mathrm{q}^2|)} \approx \frac{1}{\alpha_{s,R}(\mu^2c^2)} + \frac{21}{12\pi}\ln\Big(\frac{|\mathrm{q}^2|}{\mu^2c^2}\Big). \tag{8.11c}$$

where the values of $\alpha_{y,R}(\mu^2c^2)$, $\alpha_{w,R}(\mu^2c^2)$ and $\alpha_{s,R}(\mu^2c^2)$ are experimentally determined. The depiction of the system (8.11) in Figure 8.2 is very suggestive: the magnitudes of the $SU(3)_c$-, $SU(2)_w$- and $U(1)_y$-interactions converge and become approximately equal somewhere around $|\mathrm{q}| \sim 10^{15}\,\mathrm{GeV}/c$. The details that ensure that the three functions (8.11) really merge in one point include an increasing precision of the measurements of the "initial" values, as well as the assumption of possible new particles with masses between $m_t \sim 174.2\,\mathrm{GeV}/c^2$ and the energy where the functions (8.11) acquire the same value.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

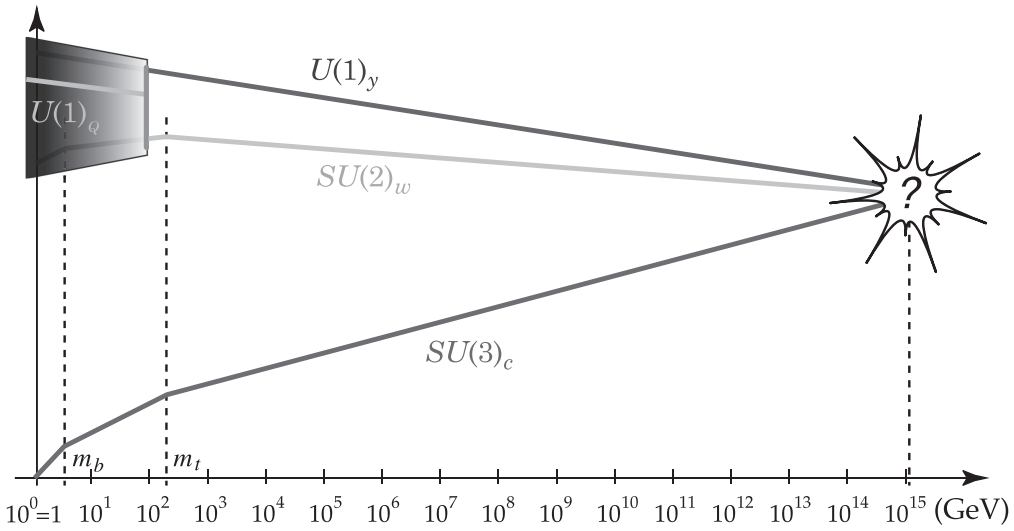

Figure 8.2 The convergence of the $SU(3)_c \times SU(2)_w \times U(1)_y$ gauge interaction strengths in the Standard Model. The slope changes indicate energy thresholds where new real quarks may be produced. The shaded area indicates the $SU(2)_w \times U(1)_y \to U(1)_Q$ phase transition.

The simplest assumption – that in this enormous span of energies nothing new exists – in fact *does not* lead to a precise merging of all three functions. In turn, in some of the possible and explored extensions of the Standard Model, this agreement is much better. One such extension is the so-called Minimally (extended) Supersymmetric Standard Model (MSSM), where this "grand desert" is populated by new particles: one superpartner for each Standard Model particle.

Of course, only concrete experiments may decide and provide the ultimate conclusion about the best model of unification of gauge interactions – as well as whether such a unification even takes place at all. As is known from even the popular literature and daily newspapers, the installations that such experiments require have in the twentieth century grown ever larger and more complex, and so are subject to both financial and political difficulties – already of international proportions. A glance into the past and the much more modest requirements of epoch-making experiments at the turn of the nineteenth into the twentieth century implies the practical impossibility of continuing one of the two pillars of experimental physics (and Rutherford's legacy): colliders (where beams of particles are accelerated and then collided, and where real collision processes are observed to happen) are becoming prohibitively expensive and complex.

The other conceptual type of experiments is based on the quantum essence of natural processes: Even if the energy in a system is insufficient for the interaction mediator in the process to be produced as a real particle, the process may nevertheless occur by exchanging virtual mediating particles. Although this significantly diminishes the probability for the observed process to happen, one then observes an enormous amount of matter (an enormous ensemble of particles) where such a process may happen, and then… waits for an unambiguous signal that the process really did happen. Until a concrete event is registered, the experiment produces *only an upper bound* for the probability for this process to happen, and cannot show if the process is in fact forbidden.

A new epoch-making advance in experimental physics will most probably require the invention of a radically new conceptual set-up of the experiment that would, in lieu of an opportunity to produce the concrete process or interaction as a real process, give a *lower bound* for the probability of this process occurring – complementary to the "waiting" experiments.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

The combination of some such new experiments with a previous type of "waiting" experiment could then narrow the limits on the probability of a process happening – which is anyway the essential goal of natural sciences [☞ Conclusion 1.1 on p. 6]. Also, if a so-obtained lower bound should surpass the independently obtained upper bound, the possibility of the process occurring would certainly be *ruled out*. To some extent, the existing experimental results from diverse installations and experiments are already being combined in such a conceptual fashion; as time passes and experimental precision grows, the available parameter space for the possible values of a considered physical quantity narrows and diminishes. However, this strategy cannot be applied to the measurement of all (20 and more, depending on the precise definition and counting) Standard Model parameters, and only a radically new type of experiments can change this.

8.2 Grand unified models

The next few sections will skim through some of the possible schemes of unification of electroweak and strong interactions.

8.2.1 The Pati–Salam $SU(4)_c \times SU(2)_L \times SU(2)_R$ model

In a series of papers [411, 410, 412] in 1973–4, Jogesh C. Pati and Abdus Salam proposed a unification scheme based on two simple ideas:

1. that "lepton-ness" is the fourth color (extending the three quark colors), and
2. that there exists a phase in which parity is an exact, i.e., restored symmetry.

These two ideas may be presented rather effectively in the form of a table:

Electroweak interaction		Chirality	$SU(4)_c$			
			$SU(3)_c$			
			r	y	b	ℓ
$SU(2)_L$	$+\frac{1}{2}$	L	u^r	u^y	u^b	ν_e^ℓ
	$-\frac{1}{2}$		d^r	d^y	d^b	$e^{-\ell}$
$SU(2)_R$	$+\frac{1}{2}$	R	u^r	u^y	u^b	ν_e^ℓ
	$-\frac{1}{2}$		d^r	d^y	d^b	$e^{-\ell}$

plus two more "families" of fundamental fermions, each with an identical structure. (8.12)

In the fully symmetric phase, the "Pati–Salam" group $SU(4)_c \times SU(2)_L \times SU(2)_R$ specifies the gauge symmetries of the model, and this certainly contains the Standard Model gauge symmetry group $SU(3)_c \times SU(2)_L \times U(1)_y$ as a subgroup. Reference [412] describes several variants of this unification, but over the subsequent years this concrete model was singled out as the most successful. The 16 fermion states in the table (8.12) are denoted typically as the

$$(\mathbf{4},\mathbf{2},\mathbf{1})_L \oplus (\mathbf{4},\mathbf{1},\mathbf{2})_R \tag{8.13}$$

representation of the $SU(4)_c \times SU(2)_L \times SU(2)_R \rtimes \mathbb{Z}_2$ group, where $\mathbb{Z}_2 = \{\mathbb{1}, P\}$ and P is the operation of parity; the symbol "$\rtimes$" denotes the semidirect product [☞ the lexicon entry, in Appendix B.1]. With respect to this complete symmetry of the Pati–Salam model, the representation (8.13) is *irreducible*, i.e., there is no proper subset of the fermions in the table (8.12), which all elements of the complete symmetry $\big(SU(4)_c \times SU(2)_L \times SU(2)_R\big) \rtimes \mathbb{Z}_2$ transform into that same subset only. In turn, since parity is a symmetry of the model, the $SU(2)_L$ and $SU(2)_R$ coupling parameters must be equal, but the $SU(4)_c$ coupling parameter is independent.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

With respect to the $SU(3)_c \times SU(2)_w \times U(1)_y \subset SU(4)_c \times SU(2)_L \times SU(2)_R \rtimes \mathbb{Z}_2$ subgroup, the representation (8.13) decomposes into

$$\big[(\mathbf{4},\mathbf{2},\mathbf{1}) \to (\mathbf{3},\mathbf{2})_{\frac{1}{3}} \oplus (\mathbf{1},\mathbf{2})_{-1}\big]_L \oplus \big[(\mathbf{4},\mathbf{1},\mathbf{2}) \to (\mathbf{3},\mathbf{1})_{\frac{4}{3}} \oplus (\mathbf{3},\mathbf{1})_{-\frac{2}{3}} \oplus (\mathbf{1},\mathbf{1})_{-2} \oplus (\mathbf{1},\mathbf{1})_0\big]_R. \quad (8.14)$$

This is the "physics-standard" notation, where the group representations are denoted by their dimensions [☞ Appendix A].[5] In particular, $(\mathbf{m},\mathbf{n},\mathbf{p})$ denotes the $SU(4)_c \times SU(2)_L \times SU(2)_R$-representation, which is the tensor product of the $\mathbf{m}$-dimensional representation of the $SU(4)_c$ group, the $\mathbf{n}$-dimensional representation of the $SU(2)_L$ group and the $\mathbf{p}$-dimensional representation of the $SU(2)_R$ group. Thus, $(\mathbf{4},\mathbf{2},\mathbf{1})$ is an $SU(4)_c$-quartet of $SU(2)_L$-pairs of quark-leptons, which decompose (8.14) into

$$(\mathbf{4},\mathbf{2},\mathbf{1}) = \begin{Bmatrix} u^r, u^y, u^b, \nu_e^\ell \\ d^r, d^y, d^b, e^{\ell-} \end{Bmatrix}_L \to \left[(\mathbf{3},\mathbf{2})_{\frac{1}{3}} = \begin{Bmatrix} u^r, u^y, u^b \\ d^r, d^y, d^b \end{Bmatrix}\right]_L \oplus \left[(\mathbf{1},\mathbf{2})_{-1} = \begin{Bmatrix} \nu_e^\ell \\ e^{\ell-} \end{Bmatrix}\right]_L, \quad (8.15a)$$

$$(\mathbf{4},\mathbf{1},\mathbf{2}) = \begin{Bmatrix} u^r, u^y, u^b, \nu_e^\ell \\ d^r, d^y, d^b, e^{\ell-} \end{Bmatrix}_R \to \begin{cases} \big[(\mathbf{3},\mathbf{1})_{\frac{4}{3}} = \{u^r, u^y, u^b\}\big]_R \oplus \big[(\mathbf{1},\mathbf{1})_{-2} = \{e^{\ell-}\}\big]_R \\ \big[(\mathbf{3},\mathbf{1})_{-\frac{2}{3}} = \{d^r, d^y, d^b\}_R\big]_R \oplus \big[(\mathbf{1},\mathbf{1})_0 = \{\nu_e^\ell\}\big]_R \end{cases}. \quad (8.15b)$$

In distinction from the left–right asymmetric interactions in the Standard Model [☞ Table 7.1 on p. 275], the extended electroweak interaction with the $SU(2)_L \times SU(2)_R$ gauge symmetry is *universal*. That is, the table (8.12) makes it clear that this model unavoidably predicts the existence of the right-handed neutrino. The right-handed neutrinos were indeed listed in Table 7.1 on p. 275, but the Standard Model does not mandate their existence. The right-handed neutrinos are invariant under the action of the Standard Model gauge symmetries and those symmetries do not link them with any other particles. In fact, all right-handed fermions in Table 7.1 on p. 275 do not partake in weak interactions and are invariant under its gauge symmetry $SU(2)_L$. In stark contrast, the left–right symmetric gauge group $SU(4)_c \times SU(2)_L \times SU(2)_R$ in the Pati–Salam model includes right-handed neutrinos in the $SU(2)_R$-doublets, extends weak interactions to left-handed particles, and thus provides the system a phase with a weak interaction that is universal (and not restricted to left-handed particles only) and where the symmetry of parity is restored.

In turn, this model then also makes it possible to describe the spontaneous breaking of the parity symmetry.

In the early 1970s, one could only suppose that there should exist a method of endowing the left- and the right-handed neutrino with masses non-symmetrically. The so-called see-saw model [☞ discussion of the relation (7.132a)–(7.132b)] was discovered only much later, and this model – the only one known – requires a mass parameter $M_\nu \gtrsim 10^{15}\,\mathrm{GeV}/c^2$. This then cannot stem from the Standard Model but may easily be the consequence of some symmetry breaking in the diagram (8.16); the critical energy of such symmetry breaking must be many orders of magnitude larger than $m_{W^\pm}c^2, m_Z c^2 \sim 10^2\,\mathrm{GeV}$. The technical method for parity breaking, so as to reproduce the experimentally observed phenomena, still remains insufficiently understood in left–right symmetric constructions such as the Pati–Salam model. In principle, one expects to be able to come up with some variant of spontaneous symmetry breaking *à la* Sections 7.1.1–7.1.2, but none of the explored models seems to be able to reproduce all experimental details.

[5] In the general case, this is not sufficiently precise, as all Lie groups except $SU(2) \cong Spin(3)$ have distinct representations of equal dimensions, but this ambiguity turns up very rarely within the examples of interest, and in those exceptional cases those distinct representations of equal dimensions are distinguished by additional decorations such as **15** and **15′** in $SU(3)$.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

The complete phase diagram – which in the 1970s was not discussed in detail – contains (at least) the five regimes ("phases")

$$\begin{array}{l|l} SU(4)_c \times SU(2)_L \times SU(2)_R \rtimes \mathbb{Z}_2 & SU(3)_c \times U(1)_{y'} \times SU(2)_L \times SU(2)_R \rtimes \mathbb{Z}_2 \\ \hline SU(4)_c \times SU(2)_L \times U(1)_R & \begin{array}{l|l} SU(3)_c \times U(1)_{y'} \times SU(2)_L \times U(1)_R & \\ \hline & \to SU(3)_c \times SU(2)_L \times U(1)_y \end{array} \end{array} \tag{8.16}$$

and the existence or absence of these (and possibly many other) regimes (and phase transitions between them) depends on the choice of the Higgs field(s) that control the symmetry-breaking process. In the display (8.16), the "vertical" phase transition (between the first two rows) is characterized by the breaking of the right-handed copy of the weak isospin gauge group, $SU(2)_R \to U(1)_R$, as well as the parity $\mathbb{Z}_2$. The "horizontal" transition is characterized by the breaking of the extended color symmetry, $SU(4)_c \to SU(3)_c \times U(1)_{y'}$. Finally, the phase transition from the regime in the right-hand side region in the second row into the regime in the last row is characterized by the breaking of the abelian (commutative) symmetries $U(1)_{y'} \times U(1)_R \to U(1)_y$. The original work on this model [412] indicated that there exists a choice of (by now reduced to eight) Higgs fields that can describe the required symmetry breaking, $SU(4)_c \times SU(2)_L \times SU(2)_R \to SU(3)_c \times SU(2)_L \times U(1)_y$.

Besides, this model also predicts the possibility of proton decay! Namely, the symmetry $SU(4)_c$ also contains transformations of any one quark into a corresponding lepton, such as $u^{r,y,b} \to \nu_e$ and $d^{r,y,b} \to e^-$. The gauge bosons that mediate such interactions, collectively named X, must violate the baryon and the lepton number, but preserve the fermion (i.e., "quark+lepton") number. Nevertheless, the decay of the proton into lighter particles – electrons, positrons, (anti)neutrinos and pions – is not possible in this model by way of exchanging only the gauge bosons, but requires also the exchange of some Higgs field(s). Owing to the conservation of the fermion number, the simplest such proton decay could be of the form $p^+ \to 3\nu_e + \pi^+$ or $p^+ \to 4\nu_e + e^+$.

The possibility that the proton is not stable was first seriously considered within this Pati–Salam model, but proton decay has not been experimentally confirmed to date.

8.2.2 The Georgi–Glashow $SU(5)$ model

Almost at the same time, Howard Georgi and Sheldon Lee Glashow suggested a competing unification model, based on the gauge group $SU(5)$ [202]. This model explicitly contains the left–right asymmetry of the Standard Model. Also, the Standard Model fermions appear within two distinct representations of the gauge group, $SU(5)$. However, since the gauge group has a single factor, unlike the Pati–Salam group, there is only one coupling parameter, and this model explains the relative ratio of the coupling parameters α_s, α_w and α_y (i.e., α_e).

The Standard Model fermions of each family are herein grouped:

$$(f_{\mathbf{10}})^{[AB]} = \begin{bmatrix} 0 & e^+ & u^r & u^y & u^b \\ & 0 & d^r & d^y & d^b \\ \hline & & 0 & \overline{u}_b & \overline{u}_y \\ \text{anti-} & & & 0 & \overline{u}_r \\ \text{symmetric} & & & & 0 \\ \text{rank-2 tensor} & & & & \end{bmatrix}_L , \qquad (\overline{f}_{\mathbf{5}^*})_A = \begin{bmatrix} e^- \\ \nu_e \\ \hline \overline{d}_r \\ \overline{d}_y \\ \overline{d}_b \end{bmatrix}_L , \qquad f_{\mathbf{1}} = (\overline{\nu}_e)_L , \tag{8.17a}$$

$$\mathbf{10} \to (\mathbf{3},\mathbf{2})_{\frac{1}{3}} \oplus (\mathbf{3}^*,\mathbf{1})_{-\frac{4}{3}} \oplus (\mathbf{1},\mathbf{1})_2 , \quad \mathbf{5}^* \to (\mathbf{3}^*,\mathbf{1})_{\frac{2}{3}} \oplus (\mathbf{1},\mathbf{2})_{-1} , \quad \mathbf{1} \to (\mathbf{1},\mathbf{1})_0 , \tag{8.17b}$$

where only the left-handed fermions are listed; clearly, $(\overline{\nu}_e)_L = \overline{\nu_{e,R}}$, so the anti-fermions of left-handed chirality represent fermions of right chirality. The indices $A, B = 1, \cdots, 5$ here count the components of the fundamental, 5-dimensional representation of the $SU(5)$ group, and $(f_{\mathbf{10}})^{[AB]}$

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

denotes the components of the antisymmetric matrix that represents the 10-dimensional representation. $\overline{f}_{\mathbf{5}^*}$ represents the conjugate fundamental, 5-dimensional representation but is here shown as a column-matrix rather than a row-matrix to save space.

The $SU(3)_c \times SU(2)_w \times U(1)_y$ gauge subgroup representations (8.17b) are identified akin to the decomposition (8.15), and were already indicated in the decomposition (8.17):

$$(\mathbf{1},\mathbf{1})_2 \leftrightarrow \left\{e^+\right\}_L, \qquad (\mathbf{3},\mathbf{2})_{\frac{1}{3}} \leftrightarrow \begin{Bmatrix} u^r, u^y, u^b \\ d^r, d^y, d^b \end{Bmatrix}_L, \qquad (\mathbf{3}^*,\mathbf{1})_{-\frac{4}{3}} \leftrightarrow \left\{\overline{u^r}, \overline{u^y}, \overline{u^b}\right\}_L, \tag{8.18a}$$

$$(\mathbf{1},\mathbf{2})_{-1} \leftrightarrow \begin{Bmatrix} \nu_e \\ e^- \end{Bmatrix}_L, \qquad (\mathbf{3}^*,\mathbf{1})_{\frac{2}{3}} \leftrightarrow \left\{\overline{d^r}, \overline{d^y}, \overline{d^b}\right\}_L, \qquad (\mathbf{1},\mathbf{1})_0 \leftrightarrow \left\{\overline{\nu_e}\right\}_L. \tag{8.18b}$$

The $SU(5)$ gauge bosons in this model contain the $SU(3)_c \times SU(2)_L \times U(1)_y$ Standard Model gauge bosons, and also six additional gauge bosons, which form an $SU(2)_L$-symmetry doublet, and an $SU(3)_c$-symmetry triplet:

$$\begin{Bmatrix} X^r \\ Y^r \end{Bmatrix}, \begin{Bmatrix} X^y \\ Y^y \end{Bmatrix}, \begin{Bmatrix} X^b \\ Y^b \end{Bmatrix}: \qquad \begin{matrix} I_3(X) = +\frac{1}{2}, & Q(X) = \frac{4}{3}, \\ I_2(Y) = -\frac{1}{2}, & Q(Y) = \frac{1}{3}. \end{matrix} \tag{8.19}$$

It is easy to find X- and Y-mediated processes in this model whereby the proton decays; for example,

$$\begin{aligned} p^+ = (u+u+d) &\to \left(u+u+(\overline{X}+e^+)\right) \to \left(u+(u+\overline{X})+e^+\right) \to \left(u+\overline{u}+e^+\right) \\ &\to \pi^0 + e^+ \quad \to \quad 2\gamma + e^+. \end{aligned} \tag{8.20}$$

Estimates of the proton lifetime then give the basic bounds for the X and Y gauge boson masses, and thus also the critical energy of the $SU(5) \to SU(3)_c \times SU(2)_L \times U(1)_y$ phase transition. Conversely, using the results $M_X, M_Y \sim 10^{15}\,\mathrm{GeV}/c^2$ from estimates such as Figure 8.2 on p. 303, it follows that the proton lifetime is $\tau_p \sim 10^{28}$–10^{29} years, which is too short: Experiments have by now raised the lower bounds to about 6.6×10^{33} years [293].

In turn, although the right-handed neutrino may be added to the fermions $f_{\mathbf{10}} \oplus \overline{f}_{\mathbf{5}^*}$, as in the decomposition (8.17), it is an $SU(5)$-invariant, i.e., neutral (chargeless) with respect to all $SU(5)$-gauge interactions. Thus, the right-handed neutrino may only have interactions of the Yukawa type (a product of two fermions and a scalar in the Lagrangian density), the coefficients of which are completely free parameters.

8.2.3 More complex models

Since the Pati–Salam $SU(4)_c \times SU(2)_L \times SU(2)_R$ model and the Georgi–Glashow $SU(5)$ unification model leave some of the Standard Model questions unanswered, it is reasonable to seek models with a gauge group that contain both the Pati–Salam and the Georgi–Glashow gauge group. It is interesting that the model built using the $SO(10)$ gauge group[6] contains both:

$$SO(10) \begin{matrix} \nearrow \\ \searrow \end{matrix} \begin{matrix} SU(4)_c \times SU(2)_L \times SU(2)_R, \\ SU(5) \times U(1)', \end{matrix} \qquad \mathbf{16}_L \begin{matrix} \nearrow \\ \searrow \end{matrix} \begin{matrix} (\mathbf{4},\mathbf{2},\mathbf{1})_L \oplus (\mathbf{4}^*,\mathbf{1},\mathbf{2})_L; \\ (\mathbf{10}_{-1})_L \oplus (\mathbf{5}^*{}_3)_L \oplus (\mathbf{1}_{-5})_L. \end{matrix} \tag{8.21}$$

In this model, all Standard Model fermions of one family – together with the right-handed neutrino – form the 16-dimensional irreducible spinor representation of the gauge group. The

[6] To be precise, this is in fact the *Spin*(10) group, the double covering of the $SO(10)$ group, so that the spinor representations are faithful, i.e., single-valued. However, in the physics literature one usually writes $SO(10)$, implicitly understanding the single-valuedness requirement.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

model is explicitly left–right symmetric and the coupling parameters α_s, α_w and α_y (i.e., α_e) all stem from a single coupling parameter of the $SO(10)$-gauge interaction. The $SO(10)$ unification thus contains the unification characteristics of both the Pati–Salam and the Georgi–Glashow models.

The number of both principal and practical puzzles of the Standard Model is thus reduced, but some of the questions still remain unanswered. Amongst them is the question: Why are there three fundamental fermion families? It would then seem reasonable to extend the gauge symmetry so as to also include a symmetry that mixes these fundamental fermion "families," the breaking of which should also explain the differences in the average masses of the fermions in the first, second and third "families." The simplest suggestion is the addition of another $SU(3)$ factor,[7] but this is evidently ad hoc, and it would be more desirable if this "familial" symmetry were a subgroup of some grand-unifying group.

The extension of the $SO(10)$ symmetry that would suffice in unifying all three fundamental fermion families into one irreducible gauge group representation, and where there exists a symmetry-breaking possibility such that precisely the three known families remain relatively light while all others (if any) acquire masses of the order $\gtrsim 10^{15}\,\mathrm{GeV}/c^2$ must be $SO(18)$. In models with such a large gauge group the number of additional particles (additional fundamental fermions, additional gauge bosons and Higgs fields) reaches many thousands, and such models are not easy to take seriously [☞ Refs. [104, 285], and the references cited therein].

Researchers of so-called GUT[8] models have explored most of the Lie groups that are sufficiently large to contain the Standard Model, but are in one way or the other minimal. In other words, since this research is mostly speculative owing to the extrapolations over enormous energies, the researchers mostly adhere to the Ockham principle, whereby the symmetry structure and the content (the fundamental particles list) of the Standard Model is extended only if this extension offers an explanation for one of the Standard Model puzzles.

Superstring theory revived interest in some of the earlier explored exotic unifying models, and foremost in a model based on the E_6 gauge symmetry group. In this model, the fermions of one family fit into the smallest (27-dimensional!) irreducible representation of the E_6 group, so each family of E_6-fermions also contains 11 completely new fermions, the absence of which from experiments must be explained separately. With the E_6-model, one often mentions a model based on the $SU(3)_c \times SU(3)_L \times SU(3)_R \subset E_6$ subgroup, dubbed "trinification."[9]

8.3 On the formalism and characteristics of scientific systems

The unification of our knowledge about Nature into a single, coherent, comprehensive and logically consistent system with as few as possible basic concepts and ideas is the leitmotiv of the foregoing exposition. The same guiding idea also permeates the remainder of this book, where the understanding of Nature so far acquired will be expanded with considerations about gravity and the geometrization of physics, aspects of a possible unification of bosons and fermions, as well as a final unification of matter, all its interactions and even spacetime.

It behooves us then to summarize the hierarchical structure that is usually referred to as a "scientific system," somewhat as a reprise of the introductory thoughts of Chapter 1, but now with the background of Chapters 2–7.

[7] This "familial" factor in the symmetry group must have a 3-dimensional representation to represent the three "families," and this 3-dimensional representation must be complex, as are the wave-functions of the fundamental fermions.

[8] GUT stands for "Grand Unified Theory."

[9] This term is indeed the amalgamation of "trinity" and "unification". Herein, trinity indicates the three $SU(3)$ factors in the gauge group; the double entendre allusion to the Holy Trinity may well be on purpose.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

8.3.1 The hierarchical structure of scientific systems

First of all, following the discussion in Section 1.1.2, by "scientific systems" one understands systems of understanding Nature that are based on iterating the cycle of observing–predicting–checking, which asymptotically improves this understanding. During this iterative process, the mathematical models and the apparatus we use to describe natural phenomena are extended and become technically more complex, and also describe Nature ever better and indicate an ever-increasing wealth of detail. No Student could fail to notice that the mathematical language that sufficed in the introductory hours of the first physics course quickly became inadequately scant, and that mastering new material in physics made it necessary to develop this mathematical language.

In retrospect, both the material mastered in other courses and that presented in Chapters 2–7 indicate the following categories of descriptive structures:

A model provides a mathematical description (surrogate) for a concrete physical system, whether this concerns a description of a concrete and simple physical system such as the pendulum or the lever, or a similarly concrete but complex system such as the Standard Model of elementary particle physics. In this description, every parameter quantifies a characteristic of the given and concrete physical system. For a more precise definition, see Procedure 11.1 on p. 416.

A theory is an axiomatic system in which a small number of physically motivated and logically consistent axioms (postulates) determines an infinite sequence of consequences that ensue with logical and mathematical rigor. Of course, we are interested in *physics* theories, of which one also expects that neither its axioms nor any of their consequences contradict Nature; the logically and mathematically incontrovertible consequences that (as yet) have not been tested are thus the predictions of the theory.

A theoretical system is a coherent and logically consistent axiomatic system that contains several distinct and otherwise independently defined and separately applicable theories.

Comment 8.5 *Here, we are primarily interested in the* ***theoretical*** *approach, hence we speak of* ***theoretical systems****. The analogous category of* ***scientific systems*** *of course includes both theoretical and experimental aspects of the system.*

During the second half of the twentieth century a subfield emerged within elementary particle physics that is usually referred to as phenomenology, and which effectively connects the ever more separated theoretical and experimental research. The scientific system then of course includes this bridging subfield.

Strictly in form, a theoretical system is indistinguishable from a theory; the difference stems from the physics application that dictates the source/motivation and justification of the axioms, as well as whether a sub-system can be applied separately. The following concrete example of two well-known theories as well as two theoretical systems containing those may serve to illustrate this.

Example 8.1 The special theory of relativity is based on two well-known postulates [☞ introductory part of Section 3.1, and in particular Definition 3.1 on p. 84, and comments], and of course the requisite mathematical apparatus that is well known from various earlier courses and was used in Chapter 3.

Similarly, quantum physics may also be introduced axiomatically. Various Authors cite different numbers of axioms: six [110], four [480], three [391] or two [29], mostly because the longer lists also contain some purely mathematical results, whereas the shorter lists presuppose the mathematical apparatus as independent (prerequisite) material [29].

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

In turn, there exist two relatively well distinguished theoretical systems, both of which include both the special theory of relativity and quantum theory:

1. relativistic quantum mechanics, and
2. relativistic quantum field theory.[10]

Both theoretical systems satisfy the above-cited requirement to contain (at least) two distinct and otherwise independently defined and separately applicable theories. The latter theoretical system is however more general: relativistic quantum field theory contains relativistic mechanics. The precise distinction between these two systems is beyond the scope of this book, but suffice it to state that an axiomatic approach to relativistic quantum field theory also requires the system of so-called Wightman axioms [572] or the Haag–Kastler alternative approach to *local quantum physics*, dubbed *algebraic quantum field theory* [254], or some other effective substitute for these.

The difference is simplest to see by comparing two standard texts, by the same Authors: Ref. [64] for relativistic quantum mechanics and Ref. [63] for relativistic quantum field theory.

The objective of distinguishing these categories of descriptive structures is to identify the important characteristics that distinguish models from theories and from theoretical systems. Consider again a concrete example: the success of Bohr's model of the hydrogen atom is oft cited as the turning point in adopting quantum physics.

Simplified, one says that classical physics cannot describe the hydrogen atom. Notice that classical physics is certainly a theoretical system, even if by classical physics one understands only classical mechanics with the additional, and simplest description of the Coulomb interaction.

We are now in a position to note the finesse (and trap) of Popper's falsifiability criterion [☞ Digression 1.1 on p. 9] – and so also of Conclusion 1.2 on p. 9: The necessity of quantizing the angular momentum indicates the falsifiability of *one concrete model* of classical physics – the classical planetary model of the atom by Rutherford, implicitly including the assumption of continuously variable angular momentum. This does not speak of the theoretical system (called classical physics) as a whole. As the classical planetary model predicts that the electrons in the orbit must lose energy via Bremsstrahlung – which of course is not the property of true atoms in Nature – one faces the format of a proof by contradiction. The logic of that type of proof indicates that at least one of the concrete assumptions (premisses) of the model must be at fault. As this includes all implicit/tacit assumptions, it is not a-priori clear that the non-classical quantization of the angular momentum is the only resolution of the disagreement between the model and Nature.

The fact that no one came up with a construction of a classical but stable planetary model of the atom[11] and many other results (Planck's black body spectrum formula, Einstein's explanation of the photoelectric effect, Compton's explanation of the effect that is now named after him) *jointly* indicate that the quantumness is convincingly indispensable in the description of Nature. That is,

[10] In practice, one always understands "quantum field theory" to be relativistic. As our present aim is to explicitly emphasize both relativity and quantumness of this theoretical system, both adjectives are explicitly stated.

[11] Note that Bohr's postulate of orbital angular momentum quantization all by itself does not necessarily preclude a purely classical explanation. For example, the complex system of Saturn's rings exhibits resonance phenomena that do provide excellent explanations for the stability of at least some of them. Similarly, the Titius–Bode rule, $\frac{1}{10}(4+3\cdot 2^n)$ for $n = -\infty, 0, 1, 2, 3\ldots$ and in units of Earth's semi-major axis, specifies the semi-major axis of each solar planet to within a small percentage except for Neptune. Neither this rule nor its generalization, Stanley Dermott's law (which then also applies to major satellites of solar planets), have a known theoretical explanation, although simulations support the belief that the regularity stems from many-body resonant phenomena [262]. It therefore simply does not follow that some as yet unknown but purely classical resonance phenomena could not in principle provide for the stability of certain select – quantized as it were – atomic orbits.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

the quantum description of Nature is the only *known* one, wherein models are *as best as known* consistent with all these and a vast many other phenomena observed in Nature.

However, the quantum description of Nature cannot possibly *refute* (or *falsify*) classical theory, since classical theory is a limiting case (i.e., an approximation [☞ Figure 8.1 on p. 298]) and so also an integral part of the quantum theory. It makes no sense to state that the whole refutes one of its integral parts or limiting cases. Quantum theory *extends* classical theory and is applicable to concrete systems where classical theory is no longer sufficiently accurate: recall the gradual transitions in the sketch in Figure 8.1 on p. 298 and the dependence of this feature on the adopted conventions of accuracy.

The analogous situation holds for the theory of relativity which of course does not refute its "non-relativistic" limiting case. The situation is analogous also with the (desired, but not yet existing) generally relativistic quantum field theory – i.e., the theory that coherently and consistently unifies both quantumness and general relativity in Nature. All so far known models that faithfully describe the various natural phenomena and aspects of Nature must be integral parts (as limiting or special cases, or as concrete applications) of this all-encompassing theoretical system.

Some of these objections to the ideas regarding falsifiability also emerge upon examining more closely the helicoidal cycle that Popper uses to describe advances in scientific knowledge [442]:

$$\boxed{\text{Problem situation}}_1 \rightarrow \boxed{\text{Tentative theory}}_1 \rightarrow \boxed{\text{Error elimination}}_1 \rightarrow \boxed{\text{Problem situation}}_2 \rightarrow \cdots \tag{8.22}$$

Here the appearance of a "problem situation" (such as an unexplained observation) triggers the creation of several competing "tentative theories" that do explain the problem situation. These are then subject to increasingly more rigorous testing (attempts at falsification), which eliminates those that turn out to be erroneous in this third step. The remaining (unfalsified) theory is then upheld until the next "problem situation" emerges and the cycle repeats.

This reminds us of the three-step iterative process "observe–model–predict" described in Section 1.1.2, which may be recast into the above format for comparison:

$$\boxed{\text{Observation}}_1 \rightarrow \boxed{\text{Model}}_1 \rightarrow \boxed{\text{Predict}}_1 \rightarrow \boxed{\text{Observe}}_2 \rightarrow \cdots \tag{8.23}$$

The following observations are immediate on comparison:

1. The outcome of an observation need not pose a "problem situation," i.e., a conflict with the previously established/trusted theory. It could be anywhere between complete confirmation and outright conflict, including indications for minor corrections. But most importantly, new observations may well imply wholly new phenomena, the qualitative *separateness* of which may not be fully understood until much later. For example, both electric and magnetic phenomena had been noticed some 24 centuries before they were systematically represented in mathematical models by Coulomb, Gauss, Ampère, Faraday, etc.
2. Models are neither theories, nor tentative theories nor conjectures, but concrete mathematical surrogates of a predefined accuracy and tolerance, and constructed *within* the framework provided by one or more pertinent theories or theoretical systems.
3. Comparisons of model predictions against Nature rarely have a binary outcome of either-true-or-false, and so can rarely lead to outright falsification of the model at hand. This is even more true of theories and theoretical systems, as discussed above and in Digression 1.1 on p. 9, but cannot be overemphasized:
   (a) relativistic physics does not falsify non-relativistic physics, but *extends* it;
   (b) quantum physics does not falsify classical physics, but *extends* it.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

Conclusion 8.2 *It is a historical fact that the contemporary description of Nature is a growing and integrated theoretical system, based on a number of postulates that is relatively small as compared to the scope and span of this description, and where the whole system as well as the candidates for additions are continually* ***filtered*** *by comparison with Nature and also by the need to form a coherent and logically consistent integration.*

In this sense is the contemporary description of Nature a growing organism.

We will return to this discussion in Section 11.5.

8.3.2 Inside indications of limitations

One of the systemically interesting characteristics of classical field theory, which includes the special theory of relativity, is that the field theory (theoretical system) indicates the limits of its own applicability.

The electrodynamics of charged particles

Start with the fact that both the formulation and the understanding of electrodynamics – as the basic example of a concrete classical field theory and with concrete application in mind – has essentially changed since the original, James Clerk Maxwell description in 1873.

In this original description, the electromagnetic field represented the deformation of aether, just as sound is a wave-like deformation of the medium through which it passes. For the aether itself, one supposed that it is at rest in Newton's absolute space and time. In this description, charged matter appears as a discontinuity in aether and in this sense is of secondary meaning.

In 1892, Hendrik A. Lorentz reformulated electrodynamics as a theory of the interactions between atomistic material particles and the all-permeating electromagnetic field, which permeates even the interior of the material particles. Lorentz initially had in mind the ions as these basic charged particles, but upon Thomson's discovery of electrons, in 1897, Lorentz's reformulation of electrodynamics focused on interactions of the electrons and the electromagnetic field. Following Lorentz's description, the charged particles are represented as little pellets of a finite size [☞ Digressions 4.1 on p. 132, 3.13 on p. 123, and 8.2 on p. 313], the electric charge of which is distributed over the surface and possibly also in the interior. Einstein's special theory of relativity – introduced as the basis for a description of electrodynamics of charged objects in motion – demands that the energy and momentum of a particle under the action of the Lorentz transformations change as components of a 4-vector, that the mass of the particle is Lorentz-invariant [☞ Chapter 3], and that they are related by equation (3.36), i.e., that the mass is the Lorentz-invariant magnitude of the 4-momentum, i.e., energy–momentum.

A way to satisfy this requirement in the Abraham–Lorentz model of an electron was never found, and all indications are that the 4-momentum, and then also the mass obtained from the relation (3.36), may be defined independently from the interactions of the electron with the electromagnetic field. From a contemporary, symmetry vantage point, the electric charge is a conserved Noether charge that corresponds to the continuous gauge symmetry (5.14), while the 4-momentum is the conserved Noether charge corresponding to spacetime translations [☞ Section 2.4.2, as well as Conclusion 9.6 on p. 329]. Since the gauge transformations (5.14) and spacetime translations are both logically and functionally independent, it follows that the mass of a particle must be independent of its electric charge.

Thus, it follows that classical electrodynamics is not complete in the sense that it does not seem capable of producing a consistent and complete model for charged material particles. As the well-known Michelson–Morley, Fizeau and other experiments imply that the concept of aether does not describe the experimental facts, it follows that one cannot go back to the original Maxwell view either, wherein material particles are "merely" discontinuities in aether.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

As noted in Digression 3.13 on p. 123, the total energy (mass) of the electric field of a point-like electron diverges. Paul A. M. Dirac, in 1938, suggested a covariant procedure for separating the finite portion of this energy.[12] This procedure, however, results in a reactive force that is proportional to the derivative of acceleration, changes the familiar expression for the Lorentz force in electrodynamics, and causes the pre-acceleration effect, where a particle starts accelerating before a force is applied [35]! It would seem to be possible to avoid the effects of pre-acceleration only if the electron were large enough so that the (changes in the) field would need enough time to permeate the particle. The current experimental bounds are orders of magnitude smaller than the so-obtained estimates; see, however, Refs. [420, 421, 336, 17, 464, 78, 79] for a rather more complex non-point-like model, which is argued to be consistent with contemporary experiments.

Digression 8.2 Digression 3.13 on p. 123 showed that the energy of the electric field of an electron – which is in the Abraham–Lorentz model to be thought of as a rotating sphere of radius r_e – equals $\alpha_e b \frac{\hbar c}{r_e}$. The value $b = \frac{1}{2}$ holds if the electric charge of the electron is uniformly distributed over the surface of the sphere, and $b = \frac{3}{5}$ if it is uniformly distributed over the interior of the sphere. At any rate, b is a constant of the order ~ 1, which is true even for more complex electron charge distributions.

If this energy – by definition necessary to bring the electric charge of the electron from infinite distances into any concrete configuration – is identified with the electron rest energy, $m_e c^2$, one obtains that the electron classical radius is

$$r_e = \alpha_e b \frac{\hbar}{m_e c} = \alpha_e \, b \, \lambda\!\!\!^{-}_e = 2.817\,940\,289\,4 \times 10^{-15}\, b\,\mathrm{m}, \tag{8.24}$$

where $\lambda\!\!\!^{-}_e = \frac{\lambda_e}{2\pi}$ is the so-called (reduced) Compton wavelength [☞ Table C.3 on p. 527] and which – up to the factor b – agrees with J. J. Thomson's estimates from collision processes. Namely, Thomson found the effective cross-section for electromagnetic radiation scattering off of non-relativistically moving electrons to be proportional to the area ${r_e}^2$, which agrees with the elementary analysis such as shown in Example 3.2 on p. 111. It follows that $b \lesssim 1$, and that b cannot be much smaller than 1. Interestingly, it is again Compton scattering – albeit at novel high-energy regimes – that may provide new information in this continuing quest [78, 79]; see also Refs. [420, 421, 336, 17, 464, 57].

In modern experiments, electrons are collided with energies of the order of 10^2 GeV, indicating that they come to a distance of about 10^{-18} m from each other – and do not show any sign of spatial structure. Down to such distances, electrons behave as point-like material particles, in full agreement with the relativistic quantum field theory, and the Gaussian distribution of the probability of finding the electron about this point, completely typical in quantum theory. The Abraham–Lorentz (and any other, classical) model of charged material particles thus does not agree with the experimental fact that $(r_e)_{\mathrm{exp.}} < 10^{-18}$ m, nor with the general theoretical result about the minimal size of the charge distribution [☞ Section 11.4]. Even the proton, which is not an elementary particle, is 2–3 orders of magnitude smaller than the classical radius of a particle with the elementary unit of electric charge.

From this one concludes that, for particles of mass m and electric charge qe, the classical radius $r_{cl} \sim \alpha_e \frac{q\hbar}{m\,c}$ and the corresponding time $t_{cl} \sim \alpha_e \frac{q\hbar}{m\,c^2}$ are the (lower) bounds of applicability of this scientific system called the classical electrodynamics of charged bodies. Notice that $\hbar$ appears

[12] This type of procedure is today referred to as "regularization" and is an integral part of renormalization computations.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press

explicitly in these bounds only owing to the definition of α_e; writing $r_{cl} \sim \frac{qe^2}{4\pi\epsilon_0 m\, c^2}$ and $t_{cl} \sim \frac{qe^2}{4\pi\epsilon_0 m\, c^3}$ instead gives these definitions a decidedly more classical appearance.

Pointillist quantum gravity

Section 1.3, and especially 1.3.3, has already discussed the joint characteristic of the combination of quantum theory and the qualitative characteristic of the theory of relativity: the existence of a minimal, Planck length. The analysis of Section 1.3.3 indicates that it suffices to endow the quantum theory with Newtonian gravity and the relativistic requirement that no material object can travel faster than the speed of light in vacuum. However, it also indicates that the Planck length, as a (lower) bound of resolution and knowledge, is not a characteristic of the quantum theory by itself, but of the theoretical system obtained by joining the quantum theory with (at least Newtonian and also with Einsteinian) gravity, and the relativistic limit on speeds $v < c$.

One thus expects this amalgamated theoretical system not to be fundamental, but to be an approximation to a more complete theoretical system. In fact, with the view that physics theories and theoretical systems only asymptotically approach their aim, the *Final Theory* is of course just a *dream*, and even an *impossible dream* – to paraphrase Refs. [553] and [549, 338], respectively. Nevertheless, the contemporary physics en route to that dream is no less real, pragmatic and successful in describing Nature as comprehensively, coherently and consistently as possible.

The (super)string theory (in fact a theoretical system) is currently the most complete candidate, and it necessarily contains quantum general-relativistic field theory, but we do not at present know enough about this complex theoretical system for a final estimate as to the measure in which this theoretical system can contain a faithful description of (our) Nature. For the most part, this uncertainty derives from the fact that many of the questions raised within and about (super)string theory have simply never before been posed. Other attempts, such as loop gravity and spacetime foam [489], as well as some more recent attempts, are insufficiently developed even just as (merely) theories of quantum gravity, and they certainly do not include matter and other interactions as (super)string theory *does*; we will return to these issues in Chapter 11.

https://doi.org/10.1017/9781009291507.012 Published online by Cambridge University Press