11

Strings: unification of all foundations of reality

The theory (i.e., theoretical *system*) of strings was already mentioned in Chapter 1 and Section 10.2, and in a very fleeting way also in the historical review in Chapter 2 – which must be supplemented before we turn to even a non-technical review of the theoretical system as well as some of the lessons from this research. The historical review will therefore introduce a few new terms, which will thereafter be clarified in the remainder of this chapter.

By now, there exist complete and pedagogically well organized texts on string theory [225, 224, 417, 483, 434, 124, 594, 398, 298, 46, 312, 251, 510], lecture notes [424, 381, 432, 170, 430], in relation to Yang–Mills gauge theories [439], recent reviews [373] and even popular books at the "guide for complete idiots" level [375, 358, 299]. This, final chapter – and the entire book – can then possibly serve only as an aperitif and prerequisite to most of this growing literature.

11.1 Strings: recycling, recycling...

Only about four decades old, the string theoretical system is based on the fundamental idea that the elementary particles – the basic building blocks of Nature and the Democritean ideal indivisibles – are not point-like. As we have concluded (with the benefit of hindsight!) already in Section 1.3.3, there exist natural limits to the tininess of elementary objects, and the "material point" is merely an idealization.[1] For students who have successfully completed a course in electrodynamics, it will be natural to regard these basic building blocks as 1-dimensional elementary objects in the next order of approximation (multipole expansion). Except, unlike a *rigid* dipole (rigid bodies being a non-relativistic idealization), strings are relativistic one-dimensional objects that possess "internal" dynamics, which essentially stems from their 1-dimensionality. However, these fundamental strings do not consist of anything "more elementary," and it is precisely the dynamics of this irreducible

[1] As was mentioned in the Preface, what exactly is identified as an elementary object is historically qualified. Chemists of the nineteenth century had rightly considered atoms as elementary; in the transition into the twentieth century, electrons and atomic nuclei appeared elementary, but it soon became evident that the nuclei consist of more elementary nucleons, and up until the last quarter of the twentieth century the list of elementary particles consisted of several leptons and a combinatorially growing list of hadrons. In the last quarter of the twentieth century, the list of elementary particles shortened to the compact Table 2.3 on p. 67, to a dozen or so elementary fermions and another dozen or so mediators of their interactions.

non-point-likeness that produces the unexpected complexity of the string theoretical system as well as many other properties with no precedent in the physics of elementary particles.

11.1.1 The original idea and application

The basic ideas (roots) of the string theoretical system date back to 1943, when Werner Heisenberg introduced the idea of the S-matrix. Namely, Heisenberg noticed that the familiar notions used in the classical description of physics (space, time, particle, etc.) need not be well defined in quantum physics, and tried to design a formalism that deals only with observable quantities [☞ quote on p. xi].[2] By definition, the S-matrix of a physical process maps precisely every incoming state into every possible outgoing state, and depends only on the positions, momenta, energies, etc., defined and measured sufficiently far from the location of all interactions. Thus designed, the incoming and the outgoing states are called *asymptotic*, and the S-matrix theory is maximally non-local in the sense that it specifies relations only between events that are sufficiently separated in spacetime.

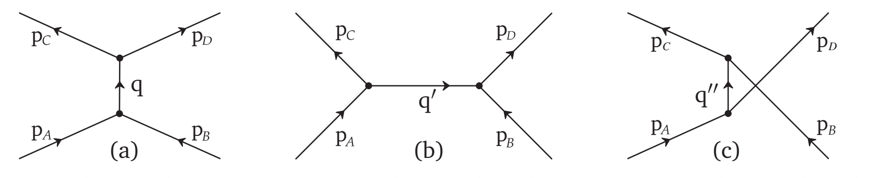
In the late 1950s and the 1960s, this approach grew into a program, the so-called "S-matrix theory," the most notable advocates of which were Stanley Mandelstam and Geoffrey Chew. Hendrik Kramers and Ralph Kronig discovered that assuming the S-matrix to be *analytic* allowed one to derive dispersion relations, which in turn imply causality between the asymptotic states even when causality is not microscopically well defined.

String theory – or, more precisely, the "dual resonant model" – was originally invented to describe hadrons, in the late 1960s. Namely, in collisions at sufficiently high energies, mesons become spatially elongated, somewhat akin to Figure 1.5 on p. 28, i.e., in display (4.101). In 1968, Gabrielle Veneziano discovered the formula (soon to be dubbed the Veneziano amplitude) that describes very well the amplitude of the effective cross-section of mesonic $A+B \to C+D$ collisions [☞ [434, Vol. 1] as well as [594]]:

$$\mathfrak{M}(\mathrm{p}_A, \mathrm{p}_B; \mathrm{p}_C, \mathrm{p}_D) \;\propto\; \frac{g^2}{\alpha'}\, I(s,t), \tag{11.1a}$$

$$I(s,t) := \int_0^1 \mathrm{d}\lambda\; \lambda^{-\alpha' s-2}(1-\lambda)^{-\alpha' t-2} = \frac{\Gamma(-\alpha' s-1)\Gamma(-\alpha' t-1)}{\Gamma(-\alpha'(s+t)-2)}, \tag{11.1b}$$

where $\Gamma(z)$ is the Euler gamma function, and the ratio $B(a,b) := \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ that appears in the Veneziano formula (11.1) is the Euler beta function. The variables s, t, u in the Veneziano amplitude (11.1) are the Mandelstam variables (3.62), which appear in the analysis of the lowest order Feynman calculus for the process $A+B \to C+D$:



$$s := -\mathrm{q}^2 = -(\mathrm{p}_A{}^2 + \mathrm{p}_B{}^2) \qquad t := -(\mathrm{q}')^2 = -(\mathrm{p}_A{}^2 - \mathrm{p}_C{}^2) \qquad u := -(\mathrm{q}'')^2 = -(\mathrm{p}_A{}^2 - \mathrm{p}_D{}^2) \tag{11.2}$$

The amplitudes for these sub-processes evidently satisfy the relations

$$\mathfrak{M}_{(a)}(p_A, p_B; p_C, p_D) = \mathfrak{M}_{(b)}(p_A, -p_C; -p_B, p_D) = \mathfrak{M}_{(c)}(p_A, -p_D; p_C, -p_D). \tag{11.3}$$

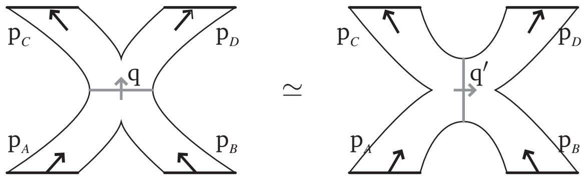
[2] One knew that protons and neutrons have finite sizes, about 10^{-15} m, and that the strength of the interaction between them at such distance was without precedent. As one after another of the attempts to model this force failed to correctly account for its peculiarities, Heisenberg believed the properties of space and time to radically change at nuclear and smaller distances.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

However, for $A+B \to C+D$ meson collisions, experiments show that the first two amplitudes are in fact *equal* – which agrees with the Veneziano formula (11.1). In other words, the first two diagrams are only two different depictions of the very same physical process, so that these two depictions are equivalent, i.e., dual to one another.

Generalizations for amplitudes of collisions where more than four incoming and outgoing mesons appear were soon discovered by Yoichiro Nambu (1968), Holger Bech Nielsen (1969) and Leonard Susskind (1969) [549]. All these results indicated that the mediating state that carries the transfer 4-momenta q, q' and q'' in the diagrams (11.2), may be represented as a linear superposition of infinitely many linear harmonic oscillators (*resonances*) with masses and frequencies that were determined by the poles of the Euler gamma function, and which are all integral multiples of a fundamental mass, i.e., frequency. This property evidently points to the possibility of interpretation of the mediating state as something filamentary, of which these linear harmonic oscillators are the Fourier modes.

This is where the original identification of mesons with open *strings* (akin to the letter "I") emerges from the so-called "dual resonant model," where *duality* refers to the equivalence in the description of the first two collision processes (11.2) using either one of the two string-diagrams:



(11.4)

It is evident that the two surfaces in the Feynman diagrams (11.4) are equal, merely specified in somewhat different parametrization and with a differing interpretation of the "mediating" state, here denoted by the gray line. In the model name, "dual resonant model," the adjective *resonant* refers to the infinite sequence of harmonic *resonances* – the Fourier modes of the mediating state – whether these are identified with a virtual string that propagates upward (time-like) in the left-hand diagram, or to the right (space-like) in the right-hand diagram. In this model, each concrete meson is identified with one of the Fourier modes of the string, whereby the incoming and the outgoing mesons are also represented as strings, fixed into the configuration of the particular Fourier mode that corresponds to the given meson.

These diagrams make it evident that strings interact by joining the end-points and splitting in two, so the two incoming open strings in the left-hand diagram (11.4) join into one, mediating, which subsequently splits into the two outgoing strings. The dual resonant model (of strings) was very popular and intensively explored in the period 1968–74, by which time certain essential properties of this model were discovered:

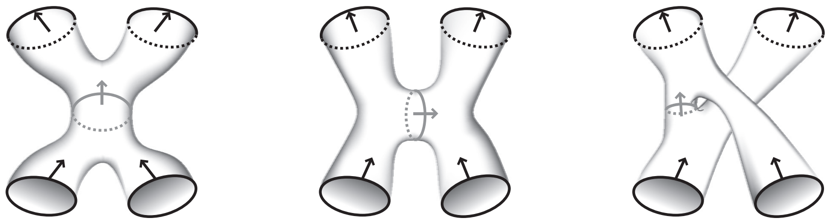
1. Open strings (akin to the letter "I") may join into closed strings (akin to the letter "O"), the Fourier modes of which have no charge and for which the effective cross-section grows with the energy. In the late 1960s, Vladimir Gribov dubbed this subset of mesons the "pomeron sector," after Isaak Pomerančuk, who proved their necessary existence in all string models.[3]
2. In the pomeron sector, the Veneziano amplitude is fully (s, t, u)-symmetric, so that

$$\mathfrak{M}(p_A, p_B; p_C, p_D) \propto \frac{g^2}{\alpha'} \big[I(s,t) + I(t,u) + I(u,s) \big], \tag{11.5}$$

[3] Geoffrey Chew and Steven Frautschi brought many of the results obtained by (then) Soviet Union physicists to the West within their "democratic" theory, in which "hadrons consist of hadrons" and have no other, "more elementary" constituent factors and which is sometimes also referred to as the "bootstrap model," alluding to a story involving Karl Friedrich Hieronymus, Freiherr von Münchhausen [65]. This reminds us of fractals – an idea that Benoit Mandelbrot would introduce in mathematics several years later, in 1975.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

which is depicted by the fact that the three string-diagrams for these three sub-processes:



(11.6)

are in fact the same surface, merely differently parametrized.

3. The pomeron sector always contains also a *massless* spin-2 Fourier mode (this particle was dubbed the *pomeron*), whereas such a hadron was never experimentally detected.
4. The consistency of the dynamics of (bosonic) strings, which must move relativistically, requires them to propagate through a flat spacetime of $25+1$ dimensions. Supersymmetric strings, which Jean-Loup Gervais and Bunji Sakita discovered in 1971 [551], consistently propagate through $(9+1)$-dimensional flat spacetime.
5. All string models without supersymmetry contain tachyons, the presence of which in the spectrum indicates an instability [☞ Digression 7.1 on p. 261]. Supersymmetry – in any model, not just string models – precludes the existence of tachyons and so stabilizes the vacuum (ground state) of the model [☞ Section 10.3.1].

The well-nigh overnight success of the quark model [☞ Sections 2.3.12 and 2.3.13] in 1974 completely suppressed the dual resonant model of hadrons and their depiction as strings, and string theory became ignored.

In the same year, 1974, Tamiaki Yoneya and independently Joël Scherk and John Schwarz noticed that string theory can be applied not as a model for hadrons, but as a model of gravity [☞ Footnote *13* on p. 13]. The pomeron (the nonexistent massless spin-2 hadron) was thus identified with the graviton, and the string *tension* (T_0) and parameter α' in the Veneziano formulae (11.1) and (11.5) were now related through the extended equality

$$\alpha' = \frac{1}{2\pi T_0 \hbar c} = \frac{G_N}{\hbar c^5}, \tag{11.7}$$

with the Newton constant, where the units are

$$[\alpha'] = \frac{T^4}{M^2 L^4} = (\text{energy})^{-2}, \qquad [T_0] = \frac{ML}{T^2} = \left[\frac{\text{energy}}{\text{length}}\right]. \tag{11.8}$$

This changes the characteristic length of strings from the characteristic size of hadrons to the Planck length:

$$\ell_S := \sqrt{\alpha'}\hbar c = \sqrt{\frac{\hbar c}{2\pi T_0}} \sim 10^{-15}\,\text{m} \quad \to \quad \ell_S = \ell_P \sim 10^{-35}\,\text{m}. \tag{11.9}$$

This suggestion remained mostly ignored, partly because of the sudden focus on the quark model, and partly owing to the fantastic requirement that spacetime in string theory would have to have 25+1 dimensions (9+1 with supersymmetry). Scherk and Schwarz supposed that the so-called (Nordstrøm)–Kałuża–Klein compactified geometry could reduce the effective spacetime dimension to 3+1.[4]

[4] The idea that space may be more than 3-dimensional, and that the additional dimensions are periodic and with too small a radius to be detected stems from Gunnar Nordstrøm, in 1914. However, this model was based on his model of gravity, which differs from the general theory of relativity, and (as determined by 1919) also from Nature. Thus, this compactification idea was mostly forgotten together with his theory of gravity, until Theodor Kałuża in 1919 and Oscar Klein in 1926 independently revived it.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

By the early 1980s, relatively few physicists were working on string theory, but even in this period of relative isolation, several significant results were derived:

1. In 1979, Daniel Friedan proved a fascinating fact about the $(1+1)$-dimensional field theory defined on the worldsheet swept out in time by strings. Requiring that quantum corrections do not renormalize the background metric of the spacetime through which the string propagates reproduces and generalizes the Einstein equations for gravity. That is, the quantum dynamics of strings implies gravity in the spacetime through which the strings move.
2. In 1981, Alexander Polyakov proved that the Hamilton action

$$S[X] = \frac{1}{4\pi\alpha'\hbar c^2} \int_\Sigma \mathrm{d}^2\xi \; g^{\alpha\beta}(\xi)\,(\partial_\alpha X^\mu)\,G_{\mu\nu}(X)\,(\partial_\beta X^\nu) \tag{11.10}$$

is classically equivalent to the Hamilton action

$$S[X] = -\frac{T_0}{c} \int_\Sigma \mathrm{d}^2\xi \; \sqrt{-\det\big[(\partial_\alpha X^\mu)\,G_{\mu\nu}(X)\,(\partial_\beta X^\nu)\big]}, \tag{11.11}$$

which Yoichiro Nambu and independently Tetsuo Goto originally formulated as the surface area of the worldsheet Σ that the string sweeps in spacetime $\mathscr{X}$. Here $X^\mu(\tau,\sigma)$ are the coordinates that indicate where in spacetime $\mathscr{X}$ is the point $\xi^\alpha = (c\tau,\sigma)$ of the string worldsheet Σ. The advantage of the Polyakov theory for the quantum theory of strings is evident: The quantization of the Hamilton action (11.11) is far more complicated (and without a generally accepted treatment) than that of the action (11.10). In turn, the geometric interpretation of equation (11.11) as the surface area of the string's worldsheet, which the Hamilton's minimal action principle minimizes, is retained by the virtue of the fact that the matrix

$$(X^*G)_{\alpha\beta} := (\partial_\alpha X^\mu)\,G_{\mu\nu}(X)\,(\partial_\beta X^\nu) \tag{11.12}$$

is the metric tensor on the worldsheet, induced ("pulled back") by the mapping $X : \Sigma \to \mathscr{X}$ from the metric tensor $G_{\mu\nu}(X)$ in spacetime.
3. Michael B. Green and John Schwarz discovered that two string models (II A and II B) are the T-dual of one another. This "T-duality" and its generalizations later proved to be one of the most important properties in the string theoretical system, by which all stringy models essentially differ from all *pointillist* models.[5]
4. In 1984, Louis Alvarez-Gaumé and Edward Witten published Ref. [12] (which had in its preprint form circulated since August 1983) with a detailed analysis of anomalies [☞ Section 7.2.3] in interactions with gravity, and with the result that the only models in $(9+1)$-dimensional spacetime where these anomalies do not destroy self-consistency are the string models II A (which is not chiral and so has no anomalies) and II B (where the anomalies cancel).

11.1.2 The string revolution

In summer of 1984, a preprint by Michael B. Green and John Schwarz started the "first string revolution," by publishing the proof [223] that Alvarez-Gaumé and Witten had omitted an important anomaly-cancellation possibility, and then demonstrated a concrete – and unexpected – mechanism (now called the Green–Schwarz mechanism) whereby anomalies cancel in the particular cases of $SO(32)$ and $E_8 \times E_8$ gauge symmetries! (Open strings with $SO(32)$ gauge symmetry were known; no string model with the $E_8 \times E_8$ gauge symmetry was then known.)

[5] To emphasize the fact that picturing elementary particles as idealized point-particles is but an approximation, in this chapter I will use the suggestive adjective *pointillist*.

Early in 1985, the papers (in circulation since late in 1984) by David J. Gross, Jeffrey A. Harvey, Emil Martinec and Ryan Rohm [247, 246, 248] were published, wherein the so-called heterotic string models were constructed: one with $SO(32)$ and another with $E_8 \times E_8$ gauge symmetry. Only a few months later, Philip Candelas, Gary Horowitz, Andy Strominger and Edward Witten showed [88], following the just-published work by Candelas with Derek Reine [90], that the $(9+1)$-dimensional spacetime of the $E_8 \times E_8$ heterotic string model may be compactified (*à la* Nordstrøm–Kałuża–Klein) on a complex Calabi–Yau 3-fold $\mathscr{Y}$,[6] so as to produce an effective model with $E_6 \times E_8$ gauge symmetry in (3+1)-dimensional spacetime and with $\frac{1}{2}\chi_E(\mathscr{Y})$ families of left-handed (chiral) fermions in the **27**-dimensional representation of the E_6 gauge group! ($\chi_E(\mathscr{Y})$ is the so-called Euler characteristic of the space $\mathscr{Y}$.) Since E_6 contains the $SU(5)$, the $SO(10)$ and the $SU(3)_c \times SU(3)_L \times SU(3)_R$ grand-unifying groups, it was clear that *there exist* string models that can contain the Standard Model of elementary particles.

Candelas soon showed that the details of the geometry of the complex Calabi–Yau 3-fold used in the compactification correlate with dynamical parameters in the Standard Model [114], which indicates that many if not all details of the Standard Model very likely may be derived from details of the geometry of so-compactified string models. This correlates many (if not all) of the physics properties of superstring models with the geometry of (generalized) spacetime!

Between spring 1984 and spring 1985, the attitude of most researchers in elementary particle physics completely changed, from totally ignoring string theory to fully focusing on constructing and exploring its models, including their compactifications.[7]

Soon, string models were constructed that at first blush had no geometric interpretation [377, 378]. Namely, in pointillist models, the configuration space is directly generated from the spacetime $\mathscr{X}$ through which these points move, and the dynamics of these material points is determined by the familiar geometry of the spacetime $\mathscr{X}$. In string models, the configuration space has a more complex structure. The Fourier spectrum of strings contains modes that propagate both in one and in the other direction around the closed string, these two classes of modes are independent and may satisfy different boundary conditions. This not only effectively doubles the configuration space but also permits constructions that are simply impossible in pointillist theories. This also requires a generalization of the usual structures in geometry – where the research is still developing☝. In 1993 Edward Witten constructed, and Jacques Distler and Shamit Kachru asymmetrically generalized the first class of models [574, 136] that interpolate between "geometric" Calabi–Yau models and "non-geometric" Landau–Ginzburg orbifold models first proposed by Cumrun Vafa. The name of these latter models indeed points to the fact that these are generalizations of the Landau–Ginzburg model as we have seen in Section 7.1.1. In these interpolating models, the "geometric" and "non-geometric" constructions appear as different phases of the same physical system defined on the worldsheet.

11.1.3 The second string revolution

It was already known in 1976 [482] that string models also include p-dimensional hypersurfaces where the open strings end if they are to satisfy Dirichlet boundary conditions [☞ Digression 11.6 on p. 415]. The integral parameter $p = 0$ here denotes a point, $p = 1$ refers to a string (filament),

---

6 The term "compactification" implies that the spacetime geometry changes $\mathscr{X} = \mathbb{R}^{1,9} \to \mathbb{R}^{1,3} \times \mathscr{Y}$, where $\mathscr{Y}$ is a compact real 6-dimensional space, and $\mathbb{R}^{1,3}$ is the (3+1)-dimensional flat spacetime. Here, $\mathscr{Y}$ is a complex 3-dimensional subspace of some better known space, typically specified by a system of algebraic equations.

7 In April 1983, I was present at a lecture by John Schwarz about strings at ICTP, in Trieste, when no one from the audience of some 300 or so active researchers in elementary particle physics asked a question after the lecture, except for Herman Nicolai who hosted the event.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

$p = 2$ to a membrane, etc. During 1989–95, Joe Polchinski showed that these so-called Dp-branes,[8] for consistency, must be treated as independent dynamical objects. They are then as fundamental as the strings from which one may have started [☞ articles [438, 433] or books [434, vol. 2] and [298]]. In this nomenclature, in the original model of open bosonic strings with $SO(32)$ gauge group, the full $(25+1)$-dimensional spacetime is a $D25$-brane, and the end-points of the open strings are restricted to this 25-dimensional space but move freely in it. Many of these various elementary p-branes are, in the string theoretical system, in fact elementary (extremely charged) black objects – a realization (recycling [☞ Footnote *13* on p. 13]) of the idea in Digression 9.5 on p. 340, just not for electrons but for these new, p-dimensional objects. This makes it clear that the name "string theory," and even "string theoretical system," is a misnomer: this theoretical system *must* include objects of various dimensions[9] [☞ Section 11.4]!

By 1995, Leonard Susskind had included Gerardus 't Hooft's holography principle in string theory, whereby the high-energy excitations of strings coincide with the thermal states of black holes and by which the fluctuations of the event horizon describe not only the degrees of freedom of the black hole itself but also of the nearby objects. That same year, Edward Witten showed that the five basic string models (open, the $SO(32)$ heterotic, the $E_8 \times E_8$ heterotic, the Type II A and the Type II B) as well as their various compactifications may be regarded as limiting cases of a more fundamental, so-called "M-theory," which he showed to also have a sixth limiting case, which contains the (otherwise unique) point-particle supergravity in $(10+1)$-dimensional spacetime [575].

Digression 11.1 M-theory extended string theory thus *incorporates* (rather than *falsifies* in a Popperian sense) point-particle field theory, and requires it to have the specific symmetries and structure of the 11-dimensional supergravity, enriched by including also specific 5-branes.

The unification of these various models into a theoretical system of strings (by now extended by various p-branes), i.e., M-theory, around 1995, is regarded as the "second string revolution," whereby the revolution of 1984 is in retrospect counted as the first. Some Authors regard the change of application of string models in 1974 (from hadrons to gravity) as the first revolution, so their counting is shifted by one from the one used herein.

Following Witten's implicit "definition" of M-theory, Tom Banks, Willy Fischler, Stephen Shenker and Leonard Susskind generalized the holography principle to the whole M-theory. Juan Maldacena then noticed [354] the following sequence of relationships:

1. Low-energy excitations near a black hole are represented by physical objects that are localized near the event horizon of the black hole.
2. In the case of extremely charged Reissner–Nordstrøm black holes, the event horizon is of the form $AdS^d \times S^{9-d}$, where AdS^d denotes the d-dimensional anti de Sitter space (9.81) and the sphere S^{9-d} carries the flux of some gauge field.
3. This latter configuration may be described by an $N = 4$ supersymmetric (and conformally symmetric) version of Yang–Mills field theory [☞ Chapter 6 and 10].

[8] The coinage "p-brane" appears in the literature, as a back-formation from *mem*brane, and where the number p denotes the number of spatial dimensions. Continuing this back-formation, the term "brane" is used even all by itself as a dimensionally non-specific collective name.

[9] This provides an opportunity for a neatly rhyming recap: *the theoretical system of strings and things*.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

This sequence of relationships was soon generalized, worked out and theoretically confirmed in detail (by Edward Witten, Steven Gubser, Igor Klebanov and Alexander Polyakov, among others [510]), and is today called the *AdS/CFT* (or, more generally, the *gravity/gauge*) *correspondence* and represents a concrete theoretical realization of the holography principle. In a roundabout way, the string theoretical system has thus returned to its original application, to the description of interactions via gauge fields.

On the other hand, soon after Witten's proposal of *M*-theory, Cumrun Vafa generalized this proposal into the so-called *F*-theory, which indicates the existence of a phase of this united theory in which the spacetime has 12 dimensions [530] – and for which the effective 12-dimensional field theory is known only in various (partially) compact variants, and for which it is not determined if it has 10+2 or 11+1 space+time dimensions (there exist arguments for both cases)☝.

11.1.4 The third string revolution

The numerologically inclined Reader must have noticed the approximate cycle of about 11 years (just like Sun-spots?) between:

1. since 1974: strings are used to describe gravity and not hadrons;
2. since 1985: the five basic string models and their compactifications;
3. since 1995–7: indirect hints of *M*- and *F*-theory as the completion of the string theoretical system, as well as the establishment of the *AdS/CFT* correspondence.

However, 2007 arrived after an unhurried percolation of several ideas that fused into the picture of *the landscape* of string theories and *the swamp* of other models [☞ the book [505], a partial critique [152], the works [531, 395], as well as the rest of this chapter for starters]. For some participants and observers, the shift in understanding the task, the purpose, and even the power of physics – with the backdrop of this landscape and swamp – represents an anti-catharsis. Namely, the very idea of the existence of an enormous and *connected* "web" of all possible string models is not new,[10] and draws its roots from the phase diagrams in grand-unified models [☞ Chapter 8], which in turn conceptually remind us of the phase diagrams in the physics of bulk materials.

The open question is, however:

1. is there a principle (such as minimization of free energy in statistical mechanics) that singles out one of all those models – hopefully such that it describes Nature just the way we observe it?
2. or is this Universe of ours selected by the fact that we – such as we are – could not even exist in some different Universe; which is the so-called "anthropic principle?"

Some regard the adoption of this anthropic principle as a revolution in understanding Nature, while others on the opposite end of a continuous palette of opinions regard this as a sign of intellectual capitulation; yet others regard it as a signal of the hopelessness of "string theory" within the science of physics [☞ paraphrasing Planck, on p. 124]. And then, there is also the infrequently adopted, but to my mind important view, best stated by Douglas Adams:

> This is rather as if you imagine a puddle waking up one morning and thinking, "This is an interesting world I find myself in – an interesting hole I find myself in – fits me rather neatly, doesn't it? In fact it fits me staggeringly well, must have been made to have me in it!" This is such a powerful idea that as the sun rises in the sky and the air heats up

[10] Early results in this direction [227, 226, 86, 87] came before their time: seven years later, their physics aspects started clearing in cases with more supersymmetry [22, 23, 24]. Two years after that, the original application of this physical process was shown to require some "isolated" models [304]; see however also recent works such as Refs. [492, 582, 290, 365] and references therein for some recent developments☝.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

> and as, gradually, the puddle gets smaller and smaller, it's still frantically hanging on to the notion that everything's going to be alright, because this world was meant to have him in it, was built to have him in it; so the moment he disappears catches him rather by surprise. I think this may be something we need to be on the watch out for. [6]

Somewhat in response to the realization of this seemingly unprecedented and complex opulence of the landscape and the swamp,[11] and the development in the vantage point (the ambivalence between the anthropic and the minimizing principle) within the string theoretical system, and partly also as a reply from the shadow of the fantastic activity around string models, there emerged critical reviews [490, 577], which are not infrequently kibitzing and criticize even just the interest in "string theory" since it is "too general to be experimentally falsified" [paraphrased, T.H.].[12] The basis of this critique stems from the fact that the string *theoretical system* has an immense (perhaps even infinite-dimensional [453]) continuous parameter space, which makes it seem impossible to refute, i.e., demonstrate that none of those choices can produce a realistic theory – or, more precisely, a concrete model that describes the observed Nature.

However, this is a very naive view of both the refutation process, as well as the logical justification of refuting. For example, the $N = 8$ supergravity in $(3+1)$-dimensional spacetime also contains a continuum of parameter choices, but since the 1980s it has been known with certainty that all the models within this theory possess the (non-realistic) unbroken parity symmetry; as of recently, we also know that all these models are most probably non-renormalizable [☞ Definition 5.1 on p. 211]. In fact, before the Green–Schwarz discovery (in 1984) of the mechanism of anomaly cancellation that was missed by the otherwise complete analysis of Alvarez-Gaumé and Witten, it was trusted that the analysis of anomalies offers a systematic characteristic of all string models (except for Type IIA and Type IIB, which in turn do not contain the Standard Model) that disqualified them all [☞ Section 11.1.2]. This definitely proves that even infinite-dimensional collections of models *can* in principle be ruled out. By virtue of the Green–Schwarz mechanism of anomaly cancellation in specific cases, the number of possible string models is significantly restricted.

In turn, even the Standard Model of elementary particles contains a continuum of parametric possibilities [☞ Chapter 7], and is reliably known to contain neither explanation nor a selection mechanism from among the possible choices of these parameters. By contrast, the last decade of research – and the imagery of the landscape and the swamp – indicates that the parameters in the string theoretical system in fact form a *discretuum* – a sufficiently dense but countable (and perhaps even finite) subset of parametric choices that is fixed among others also by a generalization of the Dirac mutual quantization of the electric and magnetic charges [☞ originally [74] and also the more recent general review [31]].

More generally and paraphrasing Refs. [505, 531, 395], there exists an immense landscape of perfectly quantum-consistent string models that contain the characteristics of Nature as we know it and which we inhabit. It is then believable that amongst those models there exists *one* that simultaneously features *all* of the characteristics of precisely *our* Nature, and with no "surplus."

[11] In reality, the plethora of theoretical models could always have been seen, but was not since many of the properties such as the dimension of spacetime or the choice of the gauge groups were taken for granted.

[12] For example, Lee Smolin [489, 490] promotes loop (quantum) gravity [☞ below] as competing with strings. Peter Woit [577] applies Pauli's denigrating "not even wrong" to "string theory" (which he openly calls a "failure") and lobbies theorists to do something else – without himself contributing to any concrete research. Bert Schroer, amid numerous historical-philosophical essays, lobbies for a very non-standard formalism that also contains a "delocalization" of point-particles [478, 479, and references quoted therein], which is "mistakenly interpreted as a string" (whatever that might mean). The critical review of Smolin's and Woit's books [490, 577] that Joe Polchinski published in the bimonthly *American Scientist* (Jan./Feb. 2007) caused a highly ramified tree of internet blog-debates that are freely accessible [☞ [436, 491, 429, 435] as well as [576], for starters].

There also exists a much larger swamp of classical, semi-classical and otherwise partially quantum-consistent models, which are however not completely quantum-consistent (and do not belong to the theoretical system of strings with its M- and F-theoretic extensions). The landscape models emerge from this swamp like islands.

> ***Conclusion 11.1*** *The (M- and F-theory extended) (super)string theoretical system is the first one in the history of fundamental physics that can predict a countability – and maybe even finiteness – of the number of perfectly quantum-consistent models with an adequate content of matter, broken P- and CP-symmetries, gravity and gauge interactions, and the first one that still maintains the hope that one of its models faithfully and completely describes* ***our*** *Nature.*

Given the fact that theoretical systems are by construction axiomatic systems that seem complex enough to be subject to Gödel's incompleteness theorem [☞ the lexicon entry, in Appendix B.1, and Appendix B.3], this feature of the string theoretical system is even surprising!

Besides, "string theory" is actually a theoretical system, which is pointless to falsify much as the *theoretical system* of classical mechanics is not falsified, nor can it be falsified in the naive Popperian sense [☞ Digression 1.1 on p. 9, Section 8.3.1 and Digression 11.2 on p. 408]. The string theoretical system contains models that are very successful in describing various natural phenomena, amongst which are also some characteristics that previously had been simply taken for granted [☞ Section 11.2]! Finally, as understood nowadays, the string theoretical system [434, Vol. 2, Figure 14.4] contains various (super)string models that are in fact limiting cases of a more fundamental theory (provisionally identified with the hints of M- or even F-theory) of which we so far know only what can be discerned from the vantage points of known special limiting cases. It is simply too early to tell.

11.1.5 Quantum gravity

The string theoretical system is by no means the only attempt at a rigorous definition and construction of both a qualitative and a quantitative description of quantum gravity [367, 322, 291, 495, 310, 67, 465, 342].

Namely, unlike Yang–Mills gauge theories, the (gauge) general theory of relativity is not renormalizable [217], nor is any field theory that includes gravity.[13] The technical part of the problem indubitably stems from the essentially nonlinear nature of the general theory of relativity [☞ Chapter 9]. Thus, a "complete theory of quantum gravity" simply (so far) does not exist (and even less existent is its quantum-consistent unification with the Standard Model). There exist the following more or less well developed candidates:

1. the (M- and F-theory extended) (super)string theoretical system,
2. loop (quantum) gravity (LQG),
3. various modifications of gravity.

The original idea in the loop (quantum) gravity approach is that the quantization procedure should be applied starting from a classical Hamilton action for gravity, which is classically equivalent to the Einstein–Hilbert action (9.38), but offers some (technical) advantage in the quantization procedure. By contrast, the Einstein equations (9.44) and the Einstein–Hilbert action (9.38) are only approximate results in the string theoretical system [☞ Section 11.1.1].

[13] So-called *perturbative gravity* is an effective description of quantum gravity where the perturbative contributions are systematically suppressed by powers of ratios of the form $E/(M_P c^2)$, where E is the characteristic energy of the considered process. In this formulation, there evidently is no chance of obtaining convergent results when the energy E approaches $M_P c^2 \sim 10^{19}\text{GeV}/c^2$, but the description becomes ever better at ever lower energies [343, and the references therein].

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

There exist more or less critical and relatively contemporary reviews of these alternative approaches (2 and 3), and the interested Reader is directed to these resources, e.g., starting with [11, 386, 385, 467, 466, 468] and the plentiful references in those works.

The alternative approach to quantizing gravity, the so-called loop (quantum) gravity, draws on the earlier *geometrodynamics* approach [310] and the use of Abhay Ashtekar's variables [21]. These were defined (in 1986) as certain contour integrals and so correspond to the (homotopy) classes of these contours. The *practice* within this approach regarding the treatment of symmetries and the subsequent constraints upon quantum states, however, radically differs from canonical (and so also Dirac, BRST, BV and BVF) quantum treatment of constraints [386, 385, 11], which is the essential building block element in the whole fundamental physics that led to the Standard Model, grand-unified models, and also the string theoretical system [☞ texts such as [554, 555, 484, 496]]. In this respect, it is not clear whether the approach to quantization in loop (quantum) gravity is in agreement with Nature as canonical quantization is known to be. Namely, the formalism of loop (quantum) gravity does not seem to detect any of the anomalies,[14] so it is not clear how this procedure could possibly be consistent with the Standard Model wherein anomalies play a crucial role.

However, the very definition (and evident possibility of identifying these contours with strings) indicates a possible connection between loop (quantum) gravity and string models. At any rate, however, loop (quantum) gravity for now does not even try seriously to unify gravity with other interactions and matter, and this approach is in this sense in a very different (and much more modest) category than the string theoretical system.

In other conceptual approaches, one postulates that on small enough distances either the Lorentz symmetry is no longer valid [☞ [273] and the recent review [545]], and perhaps even the space ceases to be a continuous topological space as one otherwise usually regards it and becomes something akin to foam [☞ Footnote *2* on p. 398], or the law of gravity changes on cosmic scales [155, and references therein], or some other characteristic of gravity and/or spacetime varies. In the spacetime foam approach, the macroscopic and well-known characteristics of spacetime (continuity, smoothness) are simply a result of averaging over an enormous ensemble of structures that are, each by itself, of a wholly different nature. For astrophysical considerations that prompt such modifications of the "ΛCDM" (Einstein gravity with a cosmological constant, Λ, and cold dark matter) model and a review of modified Newtonian dynamics (MOND) proposals, see recent reviews such as Refs. [180, 498, 156] and references therein. In loop (quantum) gravity, spacetime also emerges as a dynamical and produced structure: The space itself is produced from so-called spin networks, which in time sweep out the so-called spin foam, whereby these two a-priori independent approaches turn out to have a common point.

Somewhat more recent are approaches wherein gravity emerges from a simpler theory that included neither general coordinate invariance nor a version of Einstein's equivalence principle [486, 315]. It is also possible to apply the gauge principle to different subsets of general coordinate transformations or treat them in somewhat different ways, and so obtain differently gauged theories of gravity [451, 276].

Finally, the common characteristic of all these approaches to quantum gravity is that, so far, no feasible experiment is known that would rule out or confirm any one of them. For example, the stringy corrections to the Einstein equations are too small to be measurable except in strongly curved spacetime, as should be the case near a singularity.

---

[14] See Digression 7.2. Also, LQG fails to explain the ultimate fate of the well-known Goroff–Sagnotti two-loop divergence [217] that signals the failure of conventional renormalization within quantum gravity; see also Refs. [386, 385, 11] for a more detailed discussion of shortcomings of the LQG approach.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

11.2 The theoretical system of (super)strings

The subject matter of elementary particle physics had by the end of the twentieth century grown beyond the usual confines of a physics discipline such as, e.g., atomic physics, molecular physics or astrophysics. On one hand, the subject matter is not a single, relatively bounded domain of natural phenomena and structures, but forms a hierarchy of at least two levels of such structures[15] [☞ Table 2.5 on p. 71]:

1. hadron physics,
2. quark–lepton physics,

and a third, essentially different level, temporarily identified as

3. (*M*- and *F*-theory extended) (super)string physics, including alternative approaches such as loop (quantum) gravity, spacetime foam and other modifications of gravity.

On the other hand, (super)string theory is no longer a concrete theory of one concrete (our) reality, but a theoretical system that we hope will be able to describe our reality *also*. In the same sense, nor is classical (non-relativistic and non-quantum) physics a single, particular theory of a single, particular mechanical system, but a theoretical system applicable to a broad (continuous, in fact!) spectrum of phenomena, both natural and also completely artificial.[16]

Digression 11.2 Classical mechanics – as a very well known theoretical *system* – is perfectly applicable to force laws: $F = \alpha x^{\sqrt{17}}$, $F = \beta x^{x/x_0}$, $F = \gamma^{\frac{\arctan(x/x_0)}{\ln(1+kx)}}$, etc. From among *uncountably many* functional forms, *our* Nature chose $F = -kx + O(x^3)$ for springs, but $F = \frac{k'}{r}$ for gravity. In 3-dimensional space (and with one dimension of time), Bertrand's theorem [327, 213] guarantees that only these two force laws provide for stable orbits, but nothing – within the theoretical system of classical mechanics – prevents the existence of $F \propto \frac{1}{x+x_0}$ springs, or $F \propto -r$ gravity. Yet, no one deems classical mechanics any less "scientific" because of its inability to "predict" the correct force law from first principles.

Amusingly, extending classical mechanics by including the gauge principle permits one to derive the $\propto 1/r$ force law for gravity (in $(3+1)$-dimensional spacetime). Extending it by including some ideas about elasticity (either by postulate or as derived from the microscopic structure and interactions within materials) permits one to derive the $\propto -x$ restoring force law of springs. This supports the expectation that many of the particular but unexplained characteristics of the Standard Model will be explained by (and derived from) developments *beyond* the Standard Model itself.

The next step in the evolution of that theoretical system is provided by the theory of relativity and quantum theory; the coherent unification (but without gravity, and so without acceleration) is known as quantum field theory. Conceptually, this is a well-defined theoretical system,

[15] The delineation of these "levels" is of course *practical* but artificial; Nature is one. Just as there exist chemical processes that belong both in organic and in inorganic chemistry, so does the structure of small nuclei (those of deuterium, tritium, helium,...) belong both in nuclear and in hadron physics, and the structure and dynamics of quark bound states both in hadron and in "quark–lepton physics." Finally, the electromagnetic (and also gravitational) interaction of course appears through all this physics, from microscopic to macroscopic scales.

[16] It is not difficult to see that even a small change in the concrete value of some of the natural constants would have significant repercussions with the end result that such a World would be significantly different from ours – and, thus, very *unnatural*. Numerous humorous examples of such ilk form the scientific basis for the popular books for laymen [183].

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

in that it is known that all field theory necessary in the Standard Model of elementary particle physics [☞ Chapters 3–7] is renormalizable and has no anomalies.

The remaining step in this evolution then must include gravity, and for this the (M- and F-theory extended) (super)string theoretical system is the most successful candidate.

One hopes that the *true* fundamental *Theory of Everything* is simply finite, and that no renormalization is needed; there do exist indications that this may well be the case with (super)string models. Finally, string models as a class of theoretical constructions are still the most likely milieu for approaching the Theory of Everything. It is literally too early to judge, owing to the simple reason that the class of (M- and F-theory extended) (super)string models is by far not known sufficiently well [505] – although one knows things about string models that one would not even have thought of asking of the physics before strings.

11.2.1 General requirements

The basic characteristics of Nature, as uncovered by the physics of the twentieth century, are summarized in Table 11.1. Any theoretical system with the ambition to describe Nature must contain these characteristics as its integral properties. Alternatively, were we to wish to substitute any of the listed characteristics with something else, we would have to prove not only that the alternative equally well generates models of natural phenomena, but also that it fits equally well with all other characteristics of Nature.

Table 11.1 Characteristics of describing Nature, key properties/purpose and the resulting unifications

	Characteristic	Universal property	Unifies/describes
	Quantumness	Stabilizes atoms	Waves and particles
Gauge principle	**Special relativity**	Links symmetries, conservation laws, forces/interactions and geometry	Spacetime, energy–momentum
	General relativity		Acceleration-gravitation, mass-inertia
	Relativity of phases (of wave-functions)		(Electro-magneto) + weak, and strong interactions
	Supersymmetry*	Stabilizes vacuum	Bosons and fermions

* Supersymmetry is the only characteristic listed here that is not yet experimentally verified, but is the only (known!) universal characteristic of which the consequences include vacuum stabilization.

> ***Conclusion 11.2*** *Nature is one; the fragments of our description of Nature sooner or later must fit into a single, coherent and consistent whole.*

11.2.2 The spacetime perspective

Start with the simplest example, where an open string moves through flat space. Let σ be a coordinate along the string, so that $\sigma = 0$ and $\sigma = \ell_S$ are the end-points (and $\sigma \in [0, 2\ell_S]$ for a closed string, where we identify the ends to form a circle of circumference $2\ell_S$), and let τ be the proper time of the string.[17] If X^μ are the coordinates in the *target* spacetime $\mathscr{X}$ through which that string moves, then $(X^0(\tau,\sigma), X^1(\tau,\sigma), \dots, X^{n-1}(\tau,\sigma))$ is the n-plet of *coordinate functions* that specify where the point σ on the string, at the proper time τ, is located in the target spacetime $\mathscr{X}$, where $\dim(\mathscr{X}) = n$.

[17] Since a string is not a point and different parts of the string may move with different velocities, every point on the string has its own proper time, and we only require that the parametrization of proper time vary continuously from point to point along the string so that the pair of variables (τ, σ) provides a coordinate system covering the worldsheet swept out by the string.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

As the proper time τ passes, the string sweeps out the worldsheet on which we may introduce the general coordinates (ξ^0, ξ^1). Suppose that the string is at rest so $c\tau = \xi^0 = X^0$ and that it extends along the coordinate $X^1 = \xi^1 = \sigma$. Then, the surface of the worldsheet is obtained by the integral $\int dX^0 dX^1$. In the general case, when the string moves arbitrarily through the target spacetime, a coordinate change must include all coordinates X^μ as well as the metric tensor of the target spacetime, and the result is the Nambu–Goto action (11.11).

Geometrically, the n-tuplet of functions $X^\mu(\xi)$ provides the mapping $X : \Sigma \to \mathscr{X}$ of the worldsheet Σ (that the string sweeps out in the process of moving) in the target spacetime $\mathscr{X}$, and the Hamilton action (11.11) characterizes this mapping.

Comment 11.1 *To be precise, $X(\Sigma) \subset \mathscr{X}$ denotes the* ***image*** *of the worldsheet in the target spacetime, to be distinguished from the abstract worldsheet Σ. For example, a* ***constant*** *mapping X produces the image $X(\Sigma)$ that is a single* ***point*** *in $\mathscr{X}$.*

Namely, for every possible image of the worldsheet that connects the string in any given initial position and the string in any given final position, one can compute the action $S[X]$. The classical worldsheet is the one that minimizes the Hamilton functional $S[X]$, and provides a textbook example of the application of Hamilton's variation principle [☞ Figure 11.1]. From this vantage

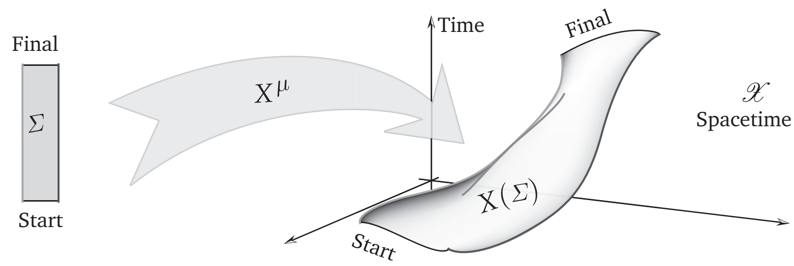


Figure 11.1 An image of the worldsheet of an open string in spacetime.

point, of primary importance is the motion of the string from its initial into its final position and the image of the worldsheet that this string sweeps out in the spacetime $X(\Sigma) \in \mathscr{X}$. The canonical quantization of the dynamics of this motion – using the Hamiltonian defined originally from the Nambu–Goto action (11.11) and subsequently from the Polyakov action (11.10) – gave the original results such as the critical dimension $\dim(\mathscr{X}) = 26$ for the ordinary string, i.e., $\dim(\mathscr{X}) = 10$ for the supersymmetric string, where the string oscillators are accompanied also by supersymmetric partners. The string end-points (i.e., the "side" edges of the surface $X(\Sigma)$) may be assigned charges. Strings interact by splitting in two and by joining ends [☞ Figure 11.2] – recall that

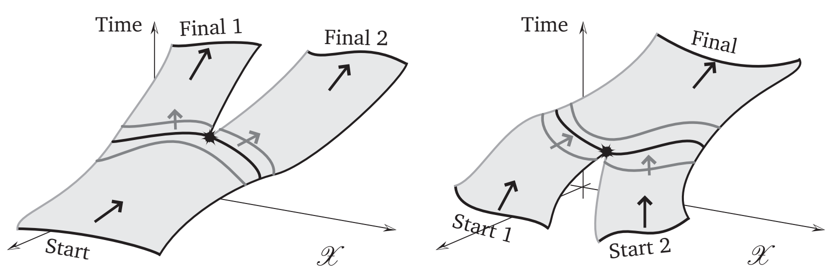


Figure 11.2 String interactions: splitting of one string into two, and joining two strings into one.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

the original inspiration for strings were mesons.[18] In mesons, the string end-points are identified as the locations of the quark and the antiquark that carry charges (isospin, electric charge, ...), while the string itself represents the continuous and two-way flux of the chromodynamic field that binds the quark and the antiquark. It follows that one end-point of a string may join with another end-point of that or another string only if all charges on one of the two end-points are of the *opposite type* from the charges on the other of the two end-points. Also, when a string splits, the two newly created end-points must have opposite charges. Thus, open strings with charged end-points represent the combination $(q, \bar{q})$, where q is the n-dimensional "charge."[19] For either of the two end-points of one string to be able to join with either of the two end-points of another, their interactions must be governed by the group $SO(n)$; quantum consistence then requires $n = 32$ [225].

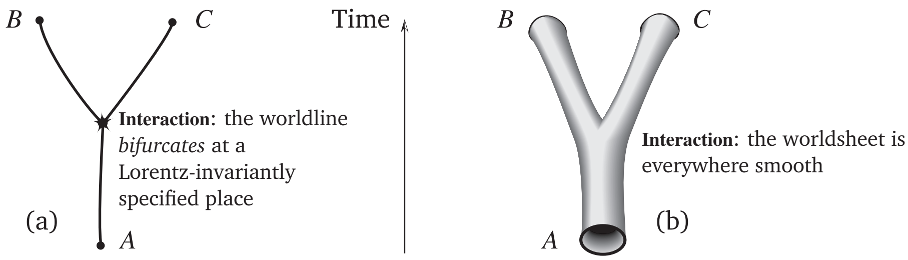


Figure 11.3 The $A \to B + C$ decay in the pointillist description (a), and in the stringy description (b).

The conceptual difference between pointillist and stringy descriptions of a simple process, e.g., the decay of a particle (3.130) $A \to B{+}C$, may be seen in Figure 11.3, for an example with a closed string. In the case of the interaction of open strings [☞ illustrations (11.4)], in certain subsets of coordinate systems there still exist special points at the edge of the string's worldsheet [☞ the cusp-like edge-points of the surfaces in the illustrations (11.4)]. However, it may be shown that this worldsheet parametrization is, via a conformal mapping, equivalent to another, where these particular points are not singled out.

Since the parametrization of the worldsheet that connects the starting and the final positions of the strings in any process cannot even in principle be measured, the measurable quantities in the stringy description must be averaged over all possible parametrizations. In doing so, not only do we have to average over all parametrizations of a fixed abstractly specified worldsheet that connects the starting with the final positions of the strings, but we must average also over all possible worldsheets that satisfy those boundary conditions. The decay process $A \to B + C$ is thus described by Feynman diagrams such as in Figure 11.3 only in the lowest approximation. The next approximation requires adding the probability amplitudes depicted by the Feynman diagrams such as in Figure 11.4. In the pointillist description of elementary particles, the two Feynman diagrams contribute differently and separately – and in fact contribute to the renormalization of different physical quantities. In the stringy variant the contributions of the two Feynman diagrams are equal, since the worldsheets of those diagrams are equivalent: one of these worldsheets may be mapped into the other by means of a continuous deformation.

[18] It is certainly not clear how to represent baryons in this naive picture, which is the additional reason for neglecting strings as a model for hadrons.

[19] In the sense that the electric charge is a 1-dimensional charge, isospin is a 2-dimensional charge, color is a 3-dimensional charge, etc.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

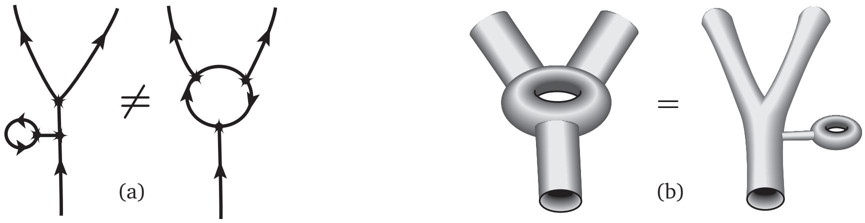


Figure 11.4 Feynman diagrams for first-order corrections to the $A \to B + C$ decay, in the pointillist description (a), and in the stringy description (b).

This engenders the intuitive impression that the string theoretical system is much better defined in the technical sense, and that the renormalization process may in fact not even be necessary.

11.2.3 The worldsheet perspective

The geometrical difference between the pointillist and the stringy depiction of physical processes [☞ Figures 11.3–11.4 on pp. 411–412] as well as the form of the action (11.10) points to an alternative perspective: the Hamilton action (11.10) evidently can be interpreted as an action for a physical system that "lives" in the (1+1)-dimensional spacetime of the *abstract* worldsheet of the string. The function $X^\mu(\xi)$ is here simply the μth field that is a scalar with respect to the (1+1)-dimensional Lorentz symmetry transformations. From this perspective, the (1+1)-dimensional worldsheet spacetime Σ is the fundamental spacetime, in which the geometry is specified by the metric tensor $g_{\alpha\beta}(\xi)$. The spacetime $\mathscr{X}$ is here simply the (abstract) *target space* in which the fields $X^\mu(\xi)$ take values and in which the metric tensor $G_{\mu\nu}(X)$ specifies how to compute the kinetic Lagrangian term for the 26-ple of fields $(X^0, \ldots, X^{25})$.

From the worldsheet perspective, the choice $\mathscr{X} = \mathbb{R}^{1,25}$ simply produces a model with 26 scalar (coordinate) fields, subject to the Polyakov action

$$\begin{aligned} S[\mathscr{X};\eta_{\mu\nu}] &= \frac{1}{4\pi\alpha'\hbar c^2}\int_\Sigma \mathrm{d}^2\xi\; g^{\alpha\beta}(\xi)\,\eta_{\mu\nu}\,(\partial_\alpha X^\mu)(\partial_\beta X^\nu) \\ &= \frac{1}{4\pi\alpha'\hbar c^2}\int_\Sigma \mathrm{d}^2\xi\; g^{\alpha\beta}(\xi)\left[(\partial_\alpha X^0)(\partial_\beta X^0) - \sum_{i=1}^{25}(\partial_\alpha X^i)(\partial_\beta X^i)\right], \end{aligned} \tag{11.13}$$

where we note that the contributions of 25 scalar functions $X^1, \ldots, X^{25}$ have the "wrong" sign of this generalized "kinetic" term,[20] which is specified by the choice of *parameters*:

$$\big[G_{\mu\nu}(X)\big]_{\text{from (11.10)}} \longmapsto \big[\eta_{\mu\nu}\big] = \mathrm{diag}[1,-1,\ldots,-1], \qquad \mu,\nu = 0,\ldots,25, \tag{11.14}$$

which in the target space $\mathscr{X} = \mathbb{R}^{1,25}$ represents the metric tensor.

[20] The sign of the whole Hamilton action depends on the choice of whether the metric tensor $g_{\alpha\beta}(\xi)$ on the worldsheet follows the "particle" or "relativist" convention – compare Chapters 3 and 9 – which is not essential here. The relative negative sign in the Lagrangian density (11.13), however, *is* essential and follows from the choice of the *signature* of the metric tensor $G_{\mu\nu}(X)$ in the target space $\mathscr{X}$.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

Digression 11.3 In the familiar application of field theory in elementary particle physics in (3+1)-dimensional spacetime, the fields in the Lagrangian densities such as (5.118), (6.23) and (7.78) represent quarks, leptons, gauge and Higgs fields. Each of these (classical!) fields takes values in a corresponding number of copies of the real or complex space, $\mathbb{R}$ or $\mathbb{C}$. The geometry – and topology – of these "target" spaces is trivial: they are *contractible*; they can be continuously contracted (by scaling) to a point, and so also contain no subspace that could *not* be continuously contracted to a point.

Digression 11.4 The essential reason for the existence of a critical dimension is precisely the non-point-like nature of strings. Owing to their spatial extension, each string itself has an infinite sequence of harmonic resonances, each of which contributes to the Hamiltonian. Each of the infinitely many such quantum oscillators contributes a non-vanishing "zero energy" [☞ the constant term in equation (10.2a)], making a formally divergent sum – but one that may be unambiguously assigned a unique, finite and definite value [259]. These contributions to the Hamiltonian are offset by the freedom of general coordinate transformations [☞ Definition 9.1 on p. 319] on the worldsheet, $\xi^\alpha(\tau,\sigma) \to \zeta^\alpha(\xi(\tau,\sigma))$. For the ground state and the observables of a quantum theory of strings to be well defined, the net "zero energy" must in fact vanish. This cancellation limits the ways in which the string can oscillate, and thereby the structure of the (generalized) spacetime probed/spanned by those oscillations [224, 434, 594, 46, 312].

In particular, the types of oscillation possible in flat spacetime limits this spacetime to have the "critical" 25+1 dimensions. Including fermionic and other types of oscillators reduces the critical dimension of the spacetime at the expense of adding structure to it. In particular, to reduce the critical dimension of this flat "target" spacetime to 9+1, one may include oscillators that generate $N=1$ supersymmetry and either $E_8 \times E_8$ or $SO(32)$ gauge symmetry, or one may include oscillators that generate $N=2$ supersymmetry in the (9+1)-dimensional spacetime.

Herein, we cannot delve into the details of such computations as existing texts do [225, 224, 434, 594, 46, 312]. Suffice it here to mention that Polchinski [434] demonstrates the existence and computes the value of the critical dimension in seven different ways; one of these is detailed accessibly by Zwiebach [594], another by Kiritsis [312]. In turn, it seems that the quantization approach in loop (quantum) gravity does not identify the anomalies of which the cancellation produces the critical dimension [386, 385], whereby the results of this quantization approach do not agree with the results of standard methods.

Amusingly, the shift in perspective, from spacetime to the worldsheet, provided inspiration to explore the analogous shift in perspective in pointillist models: from spacetime to the worldline. Evidently, instead of a field theory in (1+1)-dimensional worldsheet spacetime, here we have a field theory in (0+1)-dimensional *time* – i.e., ordinary mechanics!

Indeed, this formalism is much better understood, and the analysis ought in fact to be simpler! However, even a swift glance at the Feynman diagrams in the left-hand side of Figures 11.3 on p. 411 and 11.4 on p. 412 indicate serious difficulties: The worldline on which one is to construct the (quantum and relativistic) mechanical model *bifurcates* and so is not a Hausdorff space! In the so-designed model, one would have to define functions such as scalar fields $X^\mu(\xi)$ in the

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

expression (11.13), but now as functions of one argument, $X^\mu(\tau)$. Here, τ stands for the proper time τ, which is however not even unambiguously defined: its domain space – the worldline – is in fact not a simple line. Also, at the bifurcation points even the first derivatives with respect to the variable τ (necessary in Lagrangian dynamics) are multi-valued, and already the set-up of this approach indicates serious technical difficulties.

On the other hand, the worldsheet swept out by interacting strings is everywhere smooth (and so also all the worldsheet derivatives, not just $\partial_\alpha X^\mu(\xi)$ needed in Lagrangian dynamics) and each Hamilton action such as (11.10) is perfectly well defined – even with arbitrarily many "handles" (the right-hand diagrams in Figure 11.4 on p. 412, both have one "handle"), i.e., even for an arbitrarily high order contribution in the stringy version of the usual Feynman perturbation theory.

Digression 11.5 The shift in perspective – from spacetime, understanding this to be the "real" spacetime in which *we* live,[21] into the worldsheet spacetime – inexorably leads to the question: "Can the spacetime dimension, $n = \dim(\mathscr{X})$, be something other than $n = 4$?" In all of the "pre-stringy" development of fundamental physics through the Standard Model and beyond [☞ Chapter 8], the "obvious" dimension of spacetime, $3+1$, was taken for granted. The Nordstrøm–Kałuża–Klein model (in 1914, and 1919–26) was a small exception to this fact, but one that was forgotten owing to its initial lack of success in unifying gravity and electromagnetism – which solidified the opinion (prejudice?) that the 4-dimensionality and even uniqueness of spacetime were obvious. Also, all particle research implicitly assumed spacetime to be *flat*, open and infinitely large, $\mathscr{X} = \mathbb{R}^{1,3}$. Nontrivial geometries [☞ Chapter 9] had occupied the attention of the separate team of researchers, mostly "relativists," who for the most part did not follow the contemporary developments in elementary particle physics; in turn, neither did "particle physicists" follow the contemporary developments in the research of nontrivial solutions in the general theory of relativity.

The stringy shift in perspective irrevocably erased that chasm.

There is another property of the string theoretical model that is easiest to see from this perspective. Consider the Hamilton action (11.13) and simplify it by choosing coordinates $\xi^0 = \tau$ and $\xi^1 = \sigma$, so that $(g_{\alpha\beta}) = \begin{bmatrix} c^2 & 0 \\ 0 & -1 \end{bmatrix}$; to describe a closed string, take the coordinate σ to be periodic $\sigma \simeq \sigma + 2\pi R$. Varying the action (11.13) then produces the equations of motion

$$\partial_+\partial_- X^\mu = 0, \qquad \mu = 0, \dots, 25, \quad \partial_\pm := \tfrac{1}{2}\big[\tfrac{\partial}{\partial\sigma} \pm \tfrac{1}{c}\tfrac{\partial}{\partial\tau}\big], \tag{11.15}$$

the general solutions of which are

$$X^\mu(\tau,\sigma) = X^\mu_L(\sigma^+) + X^\mu_R(\sigma^-), \qquad \partial_+ X^\mu_R(\sigma^-) = 0 = \partial_- X^\mu_L(\sigma^+), \qquad \sigma^\pm := (c\tau \pm \sigma). \tag{11.16}$$

That is, the general solution of the D'Alembert equation in $(1+1)$-dimensional spacetime is a linear combination of two *arbitrary* functions, X^μ_L and X^μ_R, each of which depends however only on one variable, σ^+ and σ^-, respectively. Since $X^\mu_L(\sigma^+)$ remains constant when after the time $\triangle\tau > 0$

[21] This manifest *subjectivity* is the first indication that the (3+1)-dimensional spacetime in which we live is neither the only one nor is it uniquely defined in physics models. Namely, this "real" or "true" spacetime is no more "real" than the $(1+1)$-dimensional spacetime of the string worldsheet, or even the 1-dimensional worldline that a point-particle sweeps out as time passes. The spacetime in which objects (particles, strings, ...) move is typically called the target spacetime, in the sense of the mapping depicted in Figure 11.1 on p. 410, and as will become clearer in Section 11.2.4.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

the σ coordinate shifts to the *left* ($\triangle\sigma = -c\triangle\tau < 0$), the function $X^{\mu}_{L}(\sigma^{+})$ "moves" to the left; analogously, the function $X^{\mu}_{R}(\sigma^{-})$ "moves" to the right.

It is of paramount importance that the Lorentz group $Spin(1,1)$ in $(1{+}1)$-dimensional spacetime is abelian (commutative), whereby all irreducible representations are 1-dimensional. It is not hard to show that the Lorentz group does not mix $X^{\mu}_{L}(\sigma^{+})$ and $X^{\mu}_{L}(\sigma^{-})$. The only linear transformation that swaps $X^{\mu}_{L}(\sigma^{+}) \leftrightarrow X^{\mu}_{L}(\sigma^{-})$ is the discrete transformation of parity, $P : \sigma^{+} \leftrightarrow \sigma^{-}$. Thus, every string model automatically has for every scalar field $X^{\mu}(\xi)$ two independent functions $X^{\mu}_{L}(\sigma^{+})$ and $X^{\mu}_{L}(\sigma^{-})$, which may well be treated completely independently. The generalization of this phenomenon to worldsheets with arbitrary metric is only technically more complex, but conceptually remains the same, precisely owing to the signature and commutativity of the Lorentz group on the string worldsheet.

This type of "doubling" of degrees of freedom evidently does not exist in pointillist models. In turn, in models where the material points are replaced by p-dimensional objects with $p > 1$, the Lorentz group on the world-"$(p{+}1)$-volume," $Spin(1,p)$, for $p > 1$ is no longer abelian (commutative), and there is no Lorentz-invariant separation of the fields X^{μ} into two or more independent functions. The separation of fields on the string worldsheet into left- and right-moving functions is therefore a unique phenomenon in string models, and ensures the uniqueness of some of the features in string models.

Digression 11.6 The combination of these two perspectives may even offer a useful insight into a phenomenon that is harder to understand from either one of the two perspectives.

Let $X^{\hat{\mu}}(\tau,\sigma)$ be the $\hat{\mu}$th string coordinate for some fixed $\hat{\mu}$; as a function of the string proper time, τ, and coordinate σ along the string, the function $X^{\hat{\mu}}(\tau,\sigma)$ has the classical equation of motion (11.15) and one may consider two basic types of boundary conditions in the spatial coordinate σ:

$$\textbf{Dirichlet condition} \qquad X^{\hat{\mu}}(\tau,0) = x^{\hat{\mu}}_{0}, \qquad X^{\hat{\mu}}(\tau,\ell_{S}) = x^{\hat{\mu}}_{1}, \tag{11.17a}$$

$$\textbf{von Neumann condition} \qquad \partial_{\sigma}X^{\hat{\mu}}(\tau,0) = x'^{\hat{\mu}}_{0}, \quad \partial_{\sigma}X^{\hat{\mu}}(\tau,\ell_{S}) = x'^{\hat{\mu}}_{1}, \tag{11.17b}$$

where $x^{\hat{\mu}}_{0}$ and $x^{\hat{\mu}}_{1}$ are constant positions along the $\hat{\mu}$-axis, and $x'^{\hat{\mu}}_{0}$ and $x'^{\hat{\mu}}_{1}$ are dimensionless constants. The von Neumann condition at either end of the string imposes no restriction on either the position of that end-point or the velocity of its motion. Thus, string end-points with von Neumann conditions are simply free. However, imposing the Dirichlet condition to an end of a string fixes the target space position of that end, at the location $x^{\hat{\mu}}_{0}$, i.e., $x^{\hat{\mu}}_{1}$.

So, if a certain string is to satisfy the boundary conditions, say,

$$\text{von Neumann} \quad \partial_{\sigma}X^{\mu}(\tau,0) = 0, \quad \partial_{\sigma}X^{\mu}(\tau,\ell_{S}) = 0, \qquad \mu = 0,\dots,9, \tag{11.17c}$$

$$\text{Dirichlet} \quad X^{\mu}(\tau,0) = 0, \qquad X^{\mu}(\tau,\ell_{S}) = L, \qquad \mu = 10,\dots,25, \tag{11.17d}$$

the $\sigma = 0$ and $\sigma = \ell_{S}$ ends of that string are trapped on the $(9+1)$-dimensional coordinate hypersurfaces specified, respectively, by the conditions $x^{\mu}_{0} = 0$ and $x^{\mu}_{1} = L$ for $\mu = 10,\dots,25$. These two $(9+1)$-dimensional spacetime hypersurfaces are two $D9$-branes (each with 9 space-like and 1 time-like dimension).

The existence of such p-dimensional ($0 \leqslant p \leqslant 25$) objects – called "$Dp$-branes" – in string theory was already known in 1976 [482]; their dynamics was explored and emphasized only much later [434, 433, 438, 298].

---

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

11.2.4 Of models, again

It will be useful to reflect on the importance of the shift in perspective from Section 11.2.2 to Section 11.2.3. Note the common denominator in both perspectives – and with the benefit of hindsight we see that this is the case also in *all* theoretical systems of contemporary physics – the fact that scientific models have the following *canonical* geometrical content:

> ***Procedure 11.1*** *Theoretical models in general are constructed by specifying the following structural elements [☞ also the Procedure 5.1 on p. 193]:*
>
> 1. *A domain space, $\mathfrak{D}$, with local coordinates ξ.*
> 2. *A target space, $\mathfrak{T}$.*
> 3. *Maps $\varphi : \mathfrak{D} \to \mathfrak{T}$, values of which serve as local coordinates in $\mathfrak{T}$. In physics parlance, φ represents the generalized coordinates, i.e., the fields of the model.*
> 4. *The dynamical functional $S[\varphi;C] = \int_{\mathfrak{D}} \mathscr{L}(\varphi,\dot{\varphi},\dots;C)$ and boundary conditions, where C denote auxiliary parameters that specify the model.*
> 5. *Probing **currents/sources** ϑ, which are fields over the domain space $\mathfrak{D}$, chosen so that $\int_{\mathfrak{D}} \vartheta\cdot\varphi$ is invariant under the action of all symmetries and general coordinate transformations [☞ Definition 9.1 on p. 319] in the model.*
>
> ***The classical version*** *of the so-specified model is immediate: varying the action $S[\varphi;C]$ produces the equations of motion, to which one must find solutions that satisfy the given initial/boundary conditions.*

Instead of a formal proof of the identifications made in Procedure 11.1, suffice it here to consider the following examples:

Example 11.1 Non-relativistic classical mechanics of a massive point-particle

The domain space is $\mathfrak{D} = \mathbb{R}^1$ (time), the target space is $\mathfrak{T} = \mathbb{R}^3$ (space),

$$S[\vec{r}] = \int \mathrm{d}t\ L(\vec{r},\dot{\vec{r}},\dots), \qquad L(\vec{r},\dot{\vec{r}},\dots) = \tfrac{m}{2}\dot{\vec{r}}^2 - V(\vec{r}), \tag{11.18}$$

where $V(\vec{r})$ is the potential energy and $\vec{r}(t)$ the mapping $\vec{r} : (\mathfrak{D}=\mathbb{R}^1) \to (\mathfrak{T}=\mathbb{R}^3)$. Choosing the coordinates $\vec{r} = \xi^i \hat{\mathbf{e}}_i$, Hamilton's variational principle produces the Euler–Lagrange equations of motion

$$\sum_{k=0}^{\infty} (-1)^k \frac{\mathrm{d}^k}{\mathrm{d}t^k}\left[\frac{\partial L}{\partial\big(\frac{\mathrm{d}^k \xi^i}{\mathrm{d}t^k}\big)}\right] = 0, \tag{11.19}$$

which are to be solved subject to the boundary conditions for the classical solution.

Relativistic theory is conceptually the same, but with $L = -mc^2\sqrt{1-\frac{\vec{v}^2}{c^2}} - V(\vec{r})$.

Example 11.2 Non-relativistic classical mechanics of n massive point-particles

The domain space is $\mathfrak{D} = \mathbb{R}^1$ (time), the target spaces is $\mathfrak{T} = \mathbb{R}^{3n} = \oplus_{a=1}^{n}(\mathbb{R}^3)_i$,

$$S[\vec{r}_a] = \int \mathrm{d}t\ L(\vec{r}_a,\dot{\vec{r}}_a,\dots), \qquad L(\vec{r}_a,\dot{\vec{r}}_a,\dots) = \tfrac{1}{2}\sum_{a=1}^{n} m_a \dot{\vec{r}}_a^2 - V(\vec{r}_1,\cdots,\vec{r}_n), \tag{11.20}$$

where the n-tuple of 3-vectors $\{\vec{r}_a(t), a = 1,\dots,n\}$ is the mapping $\vec{r}_a : (\mathfrak{D}=\mathbb{R}^1) \to (\mathfrak{T}=\mathbb{R}^{3n})$. Choosing the coordinates $\vec{r}_a = \xi^i_a \hat{\mathbf{e}}_i$, Hamilton's variational principle produces the Euler–Lagrange equations of motion

$$\sum_{k=0}^{\infty}(-1)^k \frac{\mathrm{d}^k}{\mathrm{d}t^k}\left[\frac{\partial L}{\partial\left(\frac{\mathrm{d}^k \xi_a^i}{\mathrm{d}t^k}\right)}\right] = 0, \quad a = 1, \ldots, n, \tag{11.21}$$

which are to be solved subject to the boundary conditions for the classical solution.

Example 11.3 Classical electromagnetic field without free charges and currents
The domain space is $\mathfrak{D} = \mathbb{R}^{1,3}$ (the (3+1)-dimensional spacetime), the target space is now the quotient space [☞ Appendix A.1.1] $\mathfrak{T} = (\mathbb{R}^{1,3}/\mathbb{R}^{1,1}) \cong \mathbb{R}^{0,2}$ (the physical polarizations of all gauge fields are 4-vector potentials modulo gauge transformations),

$$S[A_\mu] = \int \mathrm{d}^4\mathrm{x}\, \mathscr{L}((\partial_\mu A_\nu), \ldots), \qquad \mathscr{L}((\partial_\mu A_\nu), \ldots) = -\tfrac{4\pi\epsilon_0}{4} F_{\mu\nu}F^{\mu\nu}, \tag{11.22}$$

where $F_{\mu\nu} := (\partial_\mu A_\nu - \partial_\nu A_\mu)$, and the 4-vector A_μ is the function that maps $A : \mathbb{R}^{1,3} \to \mathbb{R}^{1,3}$, but owing to gauge invariance all four components $A_0, \ldots, A_3$ are not independent. Imposing the Lorenz and the Coulomb gauge [☞ discussion about equation (5.91)] the temporal and longitudinal component are eliminated, which leaves the 2-dimensional space $\mathbb{R}^{0,2}$. Hamilton's variational principle produces the Euler–Lagrange equations of motion

$$\frac{\partial L}{\partial A_\mu} - \partial_\nu \frac{\partial L}{\partial(\partial_\nu A_\mu)} + \partial_\nu\partial_\rho \frac{\partial L}{\partial(\partial_\nu\partial_\rho A_\mu)} - \cdots = 0, \tag{11.23}$$

which are to be solved subject to the boundary conditions for the classical solution.

Example 11.4 Classical electromagnetic field with free charges and currents
The previous example is easily modified by adding the "probing currents/sources." These are – reading off from the Lagrangian density (5.22) – simply the *probing/test* electric charge density and current density, $\rho, \vec{\jmath}$. These should be distinguished from any charge and current density that are provided so as to produce a particular desired field. Notice that the gauge transformations (5.14a) change

$$(\rho\Phi - \vec{\jmath}\cdot\vec{A}) \;\to\; \rho\Phi - \rho\dot{\lambda} - \vec{\jmath}\cdot\vec{A} - \vec{\jmath}\cdot\vec{\nabla}\lambda = (\rho\Phi - \vec{\jmath}\cdot\vec{A}) - (\vec{\jmath}\cdot\vec{\nabla}\lambda + \rho\dot{\lambda}). \tag{11.24}$$

Under the spacetime integral, partial integrations yields

$$\delta_{\text{gauge}} \int \mathrm{d}^4\mathrm{x}\, (\rho\,\Phi - \vec{\jmath}\cdot\vec{A}) \;=\; \int \mathrm{d}^4\mathrm{x}\, (\vec{\nabla}\cdot\vec{\jmath} + \dot{\rho})\,\lambda - \int \mathrm{d}t \oint_{S^2_\infty} \mathrm{d}^2\vec{r}\, (\vec{\nabla}\cdot\vec{\jmath}\,\lambda) - \int \mathrm{d}^3\vec{r}\, [\rho\,\lambda]_{t=-\infty}^{t=+\infty}. \tag{11.25}$$

The requirement that $(\rho\Phi - \vec{\jmath}\cdot\vec{A})$ remain gauge-invariant for all *arbitrary* gauge parameter functions $\lambda(\mathrm{x})$ imposes the conditions on the probing charge and current densities: (**1**) they must vanish sufficiently fast at the boundary of the spacetime domain ($\vec{\jmath}$ at spatial infinity, ρ at time-like infinity) for the first integral to vanish after the application of Gauss's theorem, and (**2**) they must satisfy the continuity equation, $\dot{\rho} = -\vec{\nabla}\cdot\vec{\jmath}$.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

The quantum version of the model described in Procedure 11.1 may be obtained – among a historical sequence of various approaches that conceptually follow the so-called canonical quantization[22] – using the Feynman–Hibbs path-integral formalism. A proper and detailed introduction to the path-integral formalism and its applications is best deferred to the texts on the subject [459, 537, 509, 165, 123, 316, 277]. Suffice it here to provide a heuristic motivation, based on Feynman's original intuitive imagery of summing over all possible histories of a system and extending the analysis familiar from quantum mechanics.

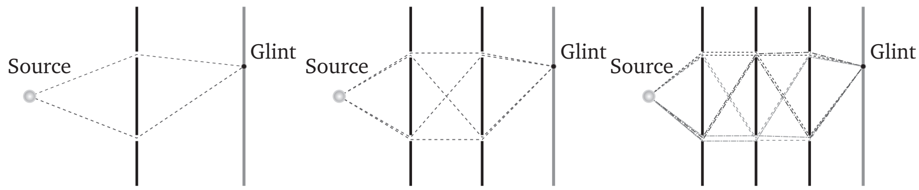


Figure 11.5 Three progressively more complicated double-slit experiments: Left, with two options, passing either through one or the other slit; middle with $2^2 = 4$ options; right, with $2^3 = 8$ options.

Consider the standard two-slit experiment, where a quantum particle is emitted from one side of the two-slitted screen and is then detected on the other side, at a specific location on a scintillating screen, as shown on the left-hand side of Figure 11.5. This arrangement is discussed in most texts of quantum mechanics, and the intensity of the glint on the scintillating screen is determined by computing the interference between the particle/wave traveling along the two distinct types of paths, one passing through the top slit, the other through the bottom one. The middle arrangement, with two consecutive two-slit experiments offers $2^2 = 4$ options: top–top, top–bottom, bottom–top and bottom–bottom. The right-most arrangement offers $2^3 = 8$ options, and so on. With n consecutive screens, each with p slits, there are p^n options for the particle passing through them, and the intensity of the glint on the scintillating screen is determined by the interference of the particle/wave traveling along all possible paths.

By letting both the number of screens and the number of slits in them grow infinitely, the collection of possible paths grows to include all possible paths that start at the source and end at the glint.[23] This is precisely as if there were no screens at all, but we expressly avoid presupposing which way a particle/wave moves from one point to the next. That in turn is precisely the situation with quantum physics, where we expressly avoid relying on classical equations of motion to determine the path of a particle between the only two points where it has been observed!

The intensity of the glint is thus determined by the total interference of the particle/wave traveling along each one from the ensemble of all possible paths from the source to the glint. Along each path, one should calculate the net phase of the particle/wave, as the difference between them determines the result of interference between particles/waves arriving along any two of the paths. This is effectively accomplished by summing phase factors $e^{i\,\text{phase}}$ over the ensemble of all possible paths. Feynman and Hibbs proved [165] that (**1**) this reasoning can be extended to all models of quantum physics, and (**2**) the correct phase-factor is in fact $\exp\{-iS[\varphi;C]/\hbar\}$, with $S[\varphi;C]$ the classical action for the system considered, as in Procedure 11.1.

[22] The contemporary BRST- and ZJBV-quantization in the Lagrangian formalism and the BFV-quantization in the Hamiltonian formalism are direct and universal generalizations of canonical quantization (as it is traditionally called), and are thus just as canonical. See Ref. [554] for pre-BRST methods, and [268, 555, 484, 496, 589, 590] for contemporary treatments.

[23] In fact, there is no a-priori reason to exclude back-tracking paths either.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

With the data provided in Procedure 11.1, one then defines the partition functional

$$Z[\vartheta;C] := \int \mathbf{D}[\varphi]\; e^{-\frac{i}{\hbar}(S[\varphi;C]+\int \mathrm{d}^4x\;\vartheta\cdot\varphi)}, \tag{11.26}$$

where $\int \mathbf{D}[\varphi]$ is the functional integral summing over all (independent and unconstrained, i.e., *free* and certainly off-shell) fields φ. This quantity turns out to be a generating function for so-called n-point correlation functions [459, 165, 123, 316, 277]

$$G(\xi_1,\xi_2,\dots) := \langle \varphi(\xi_1)\,\varphi(\xi_2)\cdots\rangle \tag{11.27}$$

$$= \frac{1}{Z[\vartheta;C]}\Big[\frac{\delta}{\delta\vartheta(\xi_1)}\frac{\delta}{\delta\vartheta(\xi_2)}\cdots Z[\vartheta;C]\Big]\Big|_{\vartheta\to 0'}\;, \tag{11.28}$$

which produce the probability amplitude for the correlation of perturbations in the field φ at the n points in the domain space, $\xi_i \in \mathfrak{D}$. They generalize the (2-point) Green function. From (11.26), one can also easily define the so-called effective action

$$S_{\text{eff.}}[\varphi_*;C'] := i\hbar\,\log\Big(\int \mathbf{D}[\varphi]\; e^{-\frac{i}{\hbar}S[(\varphi_*+\varphi);C]}\Big). \tag{11.29}$$

The fields φ_* are called "background," and provide the interpretation of φ as the (quantum) fluctuations around this background. The effective action $S_{\text{eff.}}[\varphi_*;C']$ may well be used as if a classical action, producing equations of motion for the background fields, φ_*. However, owing to the integration over all "fluctuations" in the right-hand side of the definition (11.29), $S_{\text{eff.}}$ effectively includes quantum corrections.

The concrete and technically precise use of the general framework based on (11.26)–(11.29) requires the expanse afforded by quantum field theory texts such as Refs. [459, 165, 123, 316, 277], but the above intuitive motivation, formal definitions and concepts still permit making two key observations. First, note that the definition in no way guarantees that $S_{\text{eff.}}[\varphi_*;C']$ would even resemble the original action $S[\varphi;C']$. We thus revisit Definition 5.1 on p. 211 and make it more precise:

Definition 11.1 *A quantum system is* ***renormalizable*** *if the effective action, $S_{\text{eff.}}[\varphi_*;C']$, has the same functional form and dependence of its parameters as the original action, $S[\varphi;C]$, and where the formal transformation $S[\varphi;C] \to S_{\text{eff.}}[\varphi_*;C']$ is fully described by the* ***renormalization*** *of the system parameters, $C \to C'$.*

Second, we note that the quantities such as (11.26), (11.28) and (11.29) are in practice most often calculated perturbatively, and individual contributions turn out to be calculable using Feynman diagrams – the same tools we have seen in Sections 3.3, 5.3–5.4 and 6.2. It turns out that the partition functional $Z[\vartheta;C]$ receives contributions from all possible Feynman diagrams, whereas the effective action $S_{\text{eff.}}[\varphi_*;C']$ receives contributions only from the connected diagrams. That is, contributions to the effective action are those of the partition functional, however with the contributions from disconnected Feynman diagrams subtracted [425, 586]. Both formally and in their physical meaning, these subtractions depicted by disconnected Feynman diagrams generalize the "excisions" that appear in the quantum-mechanical perturbative calculations (3.93)–(3.102).

— ❦ —

The space of all mappings $\mathfrak{T}^{\mathfrak{D}} := \{\mathfrak{D} \to \mathfrak{T}\}$ over which the integral (11.26) is to be computed – in the general case – is not the same as the configuration space. Namely, a concrete physical system is very often limited by constraints, $\chi(\varphi) = 0$, which "restrict" the mappings. The integral (11.26)

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

is defined over *free* mappings, and constraints are to be included into the dynamics by means of Lagrange multipliers $S[\varphi;C] \to S[\varphi;C] + \int_{\mathfrak{D}} \Lambda\cdot\chi(\varphi)$, so that varying the Lagrange multipliers Λ imposes the constraints $\chi(\varphi) = 0$. Among other cases, every symmetry of a system is a constraint since the symmetry transformation, S, does not change the physical state $|\Psi\rangle$, whereby it is always true that

$$\mathsf{S}|\Psi\rangle = |\Psi\rangle \quad \text{i.e.,} \quad \chi_a|\Psi\rangle = 0, \quad \text{where} \quad \chi_a := -i\frac{\partial \mathsf{S}}{\partial \epsilon^a}\Big|_{\epsilon=0} \quad \text{and} \quad \mathsf{S} = e^{i\epsilon^a \chi_a}, \tag{11.30}$$

and where ϵ^a denotes the ath symmetry transformation parameter.

Digression 11.7 In 1950–1, P. A. M. Dirac showed [132, 133] that it is essential to ensure that the dynamics of the system (in the Hamiltonian formalism, generated by the Hamiltonian) *preserves* the constraints χ_i. In classical physics, this means that all Poisson brackets $\{\chi_i, \chi_j\}$ and $\{H, \chi_i\}$ must automatically equal a linear combination of the constraints, and the model is consistent only if this can be achieved by a combination of:

1. extending the collection of all constraints, $\{\chi_1, \chi_2, \dots\}$,
2. redefining the Poisson brackets (into Dirac brackets),

whereby the (redefined or not) brackets must satisfy the Jacobi identity (6.18).

In quantum physics, the analogous situation must hold for commutators, i.e., anticommutators between spinor operators, including the generalization of the Jacobi identities (10.37). So far, the most general known procedures that ensure this are the Zinn–Justin–Batalin–Vilkovisky (BV) quantization [40, 41, 555, 484] in the Lagrangian formalism, and the Batalin–Fradkin–Vilkovisky (BFV) quantization [174, 39, 172, 36] in the Hamiltonian formalism.

Considerations and statements such as these, in this section and also in Sections 8.1.1 and 8.3, actually are not themselves part of physics (the subject matter of which is *Nature*), but of a discipline the subject matter of which is the *scientific discipline* "physics" and its structure. In the analogous situation, the discipline of which the subject matter is mathematics is called *metamathematics* [314]. Analogously, the discipline of which the subject matter is physics should be called *metaphysics*, but this name is already the standard moniker for a branch of philosophy concerned with the nature of existence and of the world.

As this discipline (about the formal structure of physics) is by its nature (just as in the description of the Procedure 11.1 and in Appendix B.3) rather mathematical, perhaps *metamatephysics* would not be too inappropriate?

11.2.5 Reconstructing the spacetime perspective

The shift in perspective from the spacetime (in Section 11.2.2) to the worldsheet (in Section 11.2.3) of course has its inverse process that reconstructs the *effective* spacetime field theory from the original worldsheet field theory.

Namely, the worldsheet Hamilton action and partition functional

$$\begin{aligned} S[X;G_{\mu\nu},\dots] &= \int_\Sigma \mathscr{L}(X^\mu,(\partial_\alpha X^\mu),\dots;\underbrace{G_{\mu\nu},\dots}_{\text{parameters}}), \\ Z[Y;G] &= \int \mathbf{D}[X]\, e^{\frac{i}{\hbar}S[X;G_{\mu\nu},\dots]+\int_\Sigma Y\cdot X}, \end{aligned} \tag{11.31}$$

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

where the parameters specify the concrete model in question. In the concrete actions (11.13) and (11.31), for example, there appears the metric tensor

$$\big[G_{\mu\nu}(X)\big]_{\text{from (11.10)}} \longmapsto \big[\eta_{\mu\nu}\big] = \text{diag}(1,-1,\dots,-1), \qquad \mu,\nu = 0,\dots,25. \tag{11.32}$$

Using the partition functional $Z[Y;G]$, we may compute how quantum fluctuations alter these specified parameters, and in 1979, Daniel Friedan showed that the condition – that these quantum fluctuations *do not alter* the chosen metric tensor – reproduces (at the lowest order in perturbative computations) the Einstein equations for the given metric tensor.

Similar results hold for all parameters in the Hamilton action (11.31): the condition for quantum stability of every parameter is an equation that in the lowest order of perturbative computation looks like a classical equation of motion for the physical quantity represented by this parameter. Thus, quantum stability of models defined on the worldsheet Σ defines the *effective* field theory in the spacetime $\mathscr{X}$, such that the classical equations of motion equal the condition for quantum stability of the original worldsheet model. Let $\{\phi^a \simeq \delta G_{\mu\nu},\dots\}$ be the collection of all parameter fluctuations in the Hamilton action (11.31).[24]

Next, construct a Hamilton action (of the second level):

$$\int_{\mathscr{X}} \mathscr{L}(\phi,(\partial_\mu\phi),\dots;\mathcal{G}_{ab},\dots), \qquad \mathscr{L} = \mathcal{G}_{ab}\,G^{\mu\nu}\,(\partial_\mu\phi^a)(\partial_\nu\phi^b) + \cdots, \tag{11.33}$$

where $\mathcal{G}_{ab}$ (and similar) parameters in the Lagrangian density (11.34) are chosen so that the classical equations of motion for the $\phi^a \simeq \delta G_{\mu\nu},\dots$ variations of the parameters in the Hamilton action of the first level precisely produce the conditions for the quantum stability of the model (11.31).

Of course, this new Hamilton action defines the quantum model

$$\begin{aligned} S[\phi;\mathcal{G}_{ab},\dots] &= \int_{\mathscr{X}} \mathscr{L}(\phi^a,(\partial_\mu\phi^a),\dots;\underbrace{\mathcal{G}_{ab},\dots}_{\text{parameters}}),\\ Z[J;\mathcal{G}] &= \int \mathbf{D}[\phi]\; e^{\frac{i}{\hbar}S[\phi;\mathcal{G}_{ab},\dots]+\int_{\mathscr{X}} J\cdot\phi}, \end{aligned} \tag{11.34}$$

where $\mathscr{F}$ is the target space, where ϕ^a take values. This quantum model is then the *effective quantum field theory* in spacetime $\mathscr{X}$ of which a part (in a realistic model, see Section 11.3) is identified with the "real" spacetime in which we live. In this model, $\phi^a \simeq \delta G_{\mu\nu},\dots$ are fields that are identified with the "real" fields such as the graviton (for $\delta G_{\mu\nu}$), the photon, the electron, the quark,... To keep the notation simple, only the graviton is explicitly written (in $\phi^a \simeq \delta G_{\mu\nu},\dots$), but each of the fields in the Standard Model may be identified as the variation of some parameter in the Hamilton action (11.31).

Evidently, the concept of generating the Hamilton action (11.34) from the previous, worldsheet Hamilton action (11.31) can be repeated: The quantum model (11.34) itself depends on parameters $\mathcal{G}_{ab},\dots$ The quantum stability of the model (11.34) produces conditions for the variations $\mathbf{\Phi}^A \sim \delta\mathcal{G}_{ab},\dots$ One then constructs a Hamilton action (of the *third level*) for the variables $\mathbf{\Phi}^A$, which is chosen so that the equations of motion for $\mathbf{\Phi}^A$ are precisely the conditions for quantum stability of the model (11.34) from the previous (second) level. This Hamilton action of the third level then defines the effective quantum field theory that "lives" in the target space of the quantum model (11.34) of the previous, second level. In principle, this iterative construction of

[24] The presentation here is drastically simplified! In practice, one must first construct the Hilbert space where the states are constructed akin to the linear harmonic oscillator, and with the creation operators from the expansion (11.38). In this Hilbert space there exist, e.g., states such as $G^{(m,n)}_{\mu\nu}(X)a^{\mu}_{m,R}a^{\nu}_{n,L}|0\rangle$, amongst which the expectation values with $m = -1 = n$ define the metric tensor $G_{\mu\nu}(X)$ on the spacetime $\mathscr{X}$ in which the coordinate fields $X^\mu(\xi)$ take values. The variables $\{\phi^a \simeq \delta G_{\mu\nu},\dots\}$ therefore parametrize the fluctuations in the Hilbert space of the worldsheet model.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

ever higher levels of effective field theories never stops; the iterative scheme may be presented formally:

$$\Big\{\,\mathfrak{D}_{(k+1)} \xrightarrow{\varphi_{(k+1)}} \mathfrak{T}_{(k+1)}\,\Big\};\quad \delta S_{(k+1)}[\varphi_{(k+1)};C_{(k+1)}]=0;\quad \delta_{\text{qu.}}C_{(k+1)}=0; \tag{11.35a}$$

$$\Big\| \qquad \nearrow\ \varphi_{(k+1)} := \delta C_{(k)} \qquad \Big\Downarrow\ \text{choice of } S_{(k+1)}[\varphi_{(k+1)};C_{(k+1)}]$$

$$\Big\{\,\mathfrak{D}_{(k)} \xrightarrow{\varphi_{(k)}} \mathfrak{T}_{(k)}\,\Big\};\qquad \underbrace{\delta S_{(k)}[\varphi_{(k)};C_{(k)}]=0;}_{\text{classical physics}}\quad \underbrace{\delta_{\text{quantum}}C_{(k)}=0;}_{\text{quantum stability}} \tag{11.35b}$$

where the scheme begins with the level where $\mathfrak{D}_{(1)} = \Sigma$ is the string worldsheet, $\mathfrak{T}_{(1)}$ the (extended) target spacetime (in the $(3+1)$-dimensional portion of which we seem to live), $\varphi_{(1)}$ are the coordinate fields immersing $\mathfrak{D}_{(1)} \to \mathfrak{T}_{(1)}$, and $C_{(1)}$ are the “coupling constants” of this first level field theory, including quantities that are in turn identified as structural characteristics of $\mathfrak{T}_{(1)}$, such as the metric tensor.

> ***Conclusion 11.3*** *String models contain an infinitely iterative hierarchy of (effective) field theories, defined by iteratively following the construction of the Hamilton action (11.34) from (11.31) [☞ scheme (11.35)].*

Of these, (at least) the first three “levels” are used routinely: the first level describes the dynamics of the (super)strings themselves (11.31), the second level describes the dynamics of the fields such as the quarks and leptons (11.34), the third level is used to explore the so-called *modular* spaces. Namely, the parameters $\mathcal{G}_{ab}$ in the second level Hamilton action, in the expressions (11.34), determine the geometry of the domain spacetime of this (second) level, and represent points in the space of possible geometries. Variations of these parameters then represent variations of these geometries and so represent local coordinates in the space of possible geometries, the so-called *modular* space. Such a modular space is then the target space in the third level and the Hamilton action in this third level then contains parameters that correspond to the structure of this modular space. In this way, the third level field theory within (super)string theories serves also as a “laboratory” for studying the structure of this modular space. It is interesting to mention that the physically motivated choice of the Zamolodchikov metric tensor on modular spaces of so-defined models coincides with the mathematically “natural” choice of the Weil–Petersson metric tensor [89], whereby the successful applications of these physical models in mathematics – amongst which some original works are collected in Refs. [85, 84] – were a fascinating surprise.

11.3 Towards realistic string models

The choice (11.13) is clearly but the simplest case, when the strings move through flat, empty and infinitely large spacetime. However, it is fairly simple to change this geometry in this model. For example, some of the scalar fields $X^\mu(\xi)$ may be required to satisfy a periodicity condition, and constantly so over the worldsheet Σ, for simplicity:

$$X^i(\xi) \simeq X^i(\xi) + 2\pi R_i, \qquad i = 4,\dots,25, \quad \forall \xi \in \Sigma. \tag{11.36}$$

As a result, the scalar fields $X^0(\xi),\dots,X^3(\xi)$ still take values in an open, flat and infinitely large space, $\mathbb{R}^{1,3}$. However, each of the scalar fields $X^4(\xi),\dots,X^{25}(\xi)$ now takes values in what is seen to be a closed and finite (*compact!*) circle of radius $R_4,\dots,R_{25}$. The shape of the target space (in which the functions $X^0(\xi),\dots,X^{25}(\xi)$ take values) has through the imposition of the conditions (11.36) turned into

$$\mathscr{X} = \mathbb{R}^{1,25} \xrightarrow{(11.36)} \mathscr{X}' = \mathbb{R}^{1,3} \times T^{22}, \quad T^{22} := S^1_{(R_4)} \times \cdots \times S^1_{(R_{25})}. \tag{11.37}$$

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

The space $\mathscr{X}'$ is compact in 22 directions, but remains non-compact in the $X^0, \dots, X^3$ directions. The conditions (11.37) represent a direct application of the Nordstrøm–Kałuża–Klein compactification.

Besides, two additional modifications may be introduced owing to the special properties of the (1+1)-dimensional worldsheet:

1. For the scalar "coordinate" fields, such as $X^4, \dots, X^{25}$, one may use the opportunity described in the discussion of the relations (11.15)–(11.16): Each of these 22 fields harbors two independent functions, upon which boundary and/or periodicity conditions may be imposed independently – and so *differently*.
2. The oscillators obtained through Fourier decomposition of functions (11.16) are bosonic and in the quantum variant correspond to creation and annihilation operators, just as with linear harmonic oscillators. Exclusively in physics defined on the (1+1)-dimensional worldsheet, every pair of bosonic creation/annihilation operators may be substituted with two pairs of *fermionic* creation/annihilation operators.[25]

The above two peculiarities of field theory in (1+1)-dimensional worldsheet spacetime (the independence of left- and right-moving modes in fields and the possibility of fermionization of bosons – and reciprocally of bosonizing fermions) makes the following construction possible:

Construction 11.1 (heterotic string) *Replace the 16 right-moving functions $X_R^{10}, \dots, X_R^{25}$ with 32 right-moving fermions, $\lambda_R^1, \dots, \lambda_R^{32}$. Impose periodicity to the left-moving functions $X_L^{10}, \dots, X_L^{25}$, so that this 16-tuple of functions $X_L^{10}, \dots, X_L^{25}$ takes values on a 16-dimensional torus that is identical with the so-called maximal torus of either the $E_8 \times E_8$ or the $D_{16} = \mathfrak{so}(32)$ algebra. The Hilbert space in such a model is built akin to the Hilbert space in Section 10.1.3, applying creation operators from the Fourier expansion following Refs. [225, 224, 434, 594],*

$$X^\mu(\tau,\sigma) = x^\mu + \tfrac{p_\mu}{p_-} c\tau + i\sqrt{\frac{\alpha'}{2}} \sum_{\substack{n=-\infty \\ n\neq 0}}^{+\infty} \left[\frac{a^\mu_{n,R}}{n} e^{-2\pi i n \varsigma^+} + \frac{a^\mu_{n,L}}{n} e^{2\pi i n \varsigma^-} \right], \tag{11.38}$$

where

$$\varsigma^\pm := (\sigma \pm c\tau)/\ell_S, \qquad \text{and} \qquad a^\mu_{-n,R} = (a^\mu_{n,R})^\dagger, \quad a^\mu_{-n,L} = (a^\mu_{n,L})^\dagger, \tag{11.39}$$

and where $p_\pm$ are the momenta corresponding to the coordinates $x^\pm := \frac{1}{\sqrt{2}}(x^0 \pm x^1)$. By choosing x^+ to be along proper time, $x^+ = \tau$, x^+ and p_+ are to be treated as simple parameters, but $x^-, x^2, \dots, x^9$ and $a^\mu_{n,R}, a^\mu_{n,L}$ and their canonically conjugate variables $p_-, p_2, \dots, p_9$ and $a^\mu_{-n,R}, a^\mu_{-n,L}$ are all the familiar quantum operators. The traditional normalization is provided by the canonical commutation relations

$$[x^-, p_-] = -i\hbar, \qquad [x^{\hat\mu}, p_{\hat\nu}] = i\hbar\delta^{\hat\mu}_{\hat\nu}, \qquad \hat\mu, \hat\nu = 2, \dots, 9; \tag{11.40a}$$

$$[a^\mu_{m,R}, a^\nu_{n,R}] = n\delta^{\mu\nu}\delta_{n,-m}, \quad [a^\mu_{m,L}, a^\nu_{n,L}] = n\delta^{\mu\nu}\delta_{n,-m}. \tag{11.40b}$$

A similar oscillator expansion also exists for $X_L^{10}, \dots, X_L^{25}$ and $\lambda_R^1, \dots, \lambda_R^{32}$. Gross, Harvey, Martinec and Rohm showed [247, 246, 248] that such a model:

[25] This transformation and its unique existence within field theory in (1+1)-dimensional spacetime was noticed in 1935–7 by Pascual Jordan, who attempted to explain a photon as a bound state of a neutrino and an antineutrino [301]; owing to the failure of this application, the basic idea of *fermionization* was itself neglected until its successful recycling [☞ Footnote *13* on p. 13] within string theory.

1. *effectively describes the propagation of strings through a* $(9{+}1)$*-dimensional, flat and infinitely large spacetime,* $\mathbb{R}^{1,9}$*, parametrized by the values of the scalar fields* $X^0(\xi), \dots, X^9(\xi)$*;*
2. *the geometric and physical meaning of* $x^0, \dots, x^9$ *is that they are the center of mass coordinates of the string;*
3. *with a suitable choice of the relative coefficients in the Hamilton action, the system exhibits a supersymmetry owing to the presence of fermionic modes that replaced the bosonic functions* $X_R^{10}, \dots, X_R^{25}$*;*
4. *with a suitable choice of the radii* $R_{10}, \dots, R_{25}$ *and angles between the coordinates* $X^{10}, \dots, X^{25}$*, the system also has either the* $E_8 \times E_8$ *or the* $SO(32)$ *gauge symmetry.*

These two specifically stringy constructions are called *heterotic* strings and proffer the possibility to construct models that contain (much more than) enough gauge symmetry and matter to describe the "real world" [247, 246, 248].

This telegraphic synopsis is a far cry from describing the details of the construction of these string models, whereby the construction certainly appears to be rather ad hoc. The technical details and consistency conditions in these constructions are, however, very rigorous; for example, these conditions single out the gauge group to be either $E_8 \times E_8$ or $SO(32)$, and the number of "flat and infinitely large" spacetime dimensions to be precisely $9{+}1$ and only if the model is supersymmetric.

11.3.1 Partially compact topology and geometry

Somewhat in the manner of the Nordstrøm–Kałuża–Klein compactification (11.37), Philip Candelas, Gary Horowitz, Andy Strominger and Edward Witten showed that by substituting the spacetime geometry

$$\mathbb{R}^{1,9} \longrightarrow \mathbb{R}^{1,3} \times \mathscr{Y} \tag{11.41}$$

in the $E_8 \times E_8$ heterotic string model, the system remains minimally (simply) supersymmetric precisely if the real 6-dimensional space $\mathscr{Y}$ is chosen to be in fact a complex 3-dimensional, compact and so-called Calabi–Yau space,[26] a.k.a., 3-fold. The hallmark feature of such spaces is that they admit a metric tensor $g_{i\bar{\imath}}$ for which:

1. $g_{i\bar{\imath}}$ is Kähler, i.e., $g_{i\bar{\imath}} = \frac{\partial}{\partial z^i}\frac{\partial}{\partial z^{\bar{\imath}}} K(z,\bar{z})$, where $K(z,\bar{z})$ is the Kähler potential;
2. the Ricci tensor computed from this metric is a total derivative, so $\oint_S \mathrm{d}z \cdot R \cdot \mathrm{d}\bar{z} = 0$ for every closed real 2-dimensional (complex 1-dimensional) surface $S \in \mathscr{Y}$.

Here (z^1, z^2, z^3) are complex local coordinates for $\mathscr{Y}$, and $z^{\bar{\imath}} = (z^i)^*$. Since the space $\mathbb{R}^{1,3} \times \mathscr{Y}$ is compact in the $\mathscr{Y}$-directions, such constructions are referred to as *Calabi–Yau compactifications*. Note that, unlike the original Nordstrøm–Kałuża–Klein compactification, here the metric tensor components in the directions of the compact space $\mathscr{Y}$ do not produce any gauge fields, as Calabi–Yau 3-folds have no isometries.

Without delving into the details of such constructions (to which end the interested Reader is directed to the book [279] and the references therein), may it suffice here to mention that the complex 3-dimensional compact Calabi–Yau spaces have two variable characteristic numbers, denoted $h^{1,1}$ and $h^{2,1}$, and that Calabi–Yau compactifications of the $E_8 \times E_8$ heterotic string models produce:

[26] Eugenio Calabi's conjecture, that the necessary and sufficient criterion for a complex 3-dimensional, compact space to admit a Käbler metric is that its first Chern class should vanish, was proven in 1974 by Shing-Tung Yau, for which he was awarded the Fields Medal. For a detailed history of both the related mathematical discoveries as well as their applications in physics and especially in (super)string theory, see Ref. [584].

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

1. an effectively (3+1)-dimensional, flat and infinitely large spacetime, $\mathbb{R}^{1,3}$;
2. minimal (simple) supersymmetry, as described in Section 10.3.2;
3. the gauge symmetry group is reduced to $E_6 \times E_8$;
4. matter fields, in the following collections:
   (a) $h^{2,1}$ copies of the **27**-representation of the group E_6,
   (b) $h^{1,1}$ copies of the $\overline{\mathbf{27}}$-representation of the group E_6;
5. a connection between the otherwise arbitrary Standard Model parameters, such as the Yukawa coupling parameters h_e, h_u, h_d, h_ν [☞ Section 7.3.1] and the geometry of the selected Calabi–Yau space $\mathscr{Y}$.

Every **27**-representation of the (compactification-reduced) gauge group E_6 contains one family of Standard Model fundamental fermions with the usual $SU(3)_c \times SU(2)_w \times U(1)_Q \subset E_6$ charges, and the $\overline{\mathbf{27}}$-representation contains the same particles but with wrong (opposite) charges. Ideally, one would like to construct a Calabi–Yau space with $h^{2,1} = 3$ and $h^{1,1} = 0$, or the other way around.[27] Besides, if the given Calabi–Yau space is not simply connected, it is possible to establish a flux of "background" gauge field along a closed contour that cannot be continuously contracted to a point. Such a non-contractible closed-contour integral of such a flux effectively serves as a Higgs field: It breaks the gauge symmetry and can "pair" fields from the **27**-representation with fields from the $\overline{\mathbf{27}}$-representation and provide them with a mass of the order of $10^{17-19}\,\text{GeV}/c^2$. For further details about constructing Calabi–Yau spaces and analyzing the models obtained by compactifying on Calabi–Yau spaces, the interested Reader is directed to search the contemporary literature (at `www.arXiv.org`), perhaps with some help from the by now two decades old Ref. [279] for starters.

11.3.2 Mirror symmetry

The analysis of the application of Calabi–Yau spaces in compactification of string models discovered the phenomenon that for every model with $h^{2,1}$ "families" and $h^{1,1}$ anti-"families" one may construct a "mirror-dual" model in which the number of "families" and anti-"families" is flipped (Brian Greene and Ronen Plesser, 1990 [236]):

$$(\mathscr{Y}, \mathscr{Y}') : \qquad h^{2,1}(\mathscr{Y}) = h^{1,1}(\mathscr{Y}'), \quad h^{1,1}(\mathscr{Y}) = h^{2,1}(\mathscr{Y}'). \tag{11.42}$$

It was soon proved [51, 83] that the phenomenon is rather typical, which provided support for a research field that may rightly be called *experimental mathematics*. Namely, between 1984 and 2002, the catalogue of constructions grew from a handful to nearly half a billion [☞ review [321] and references therein], and the statistically significant tendencies in this collection acquire significant probabilities of being systemic, and then are well worth exploring as candidates for mathematically rigorous theorems. Besides, the insight into the physical qualities of the models in which these tendencies are noted may provide an argumentation that is fully alien to the mathematical tradition, whence these tendencies may appear surprising or even "magical" from the mathematicians' vantage point.

One such example is precisely the "mirror duality," where the identifications are in fact not only on the level of numerical characteristics (11.42), but also among certain physically motivated algebraic structures that correspond to Yukawa interactions (7.133f). Using this physics insight, mirror duality may be used to compute certain mathematical characteristics of Calabi–Yau spaces [85, 84, 49, 50], to which end the "purely mathematical" methods are still not known, and which at first seemed fantastic and unbelievable.

However, intrigued by the manifest computational efficiency, mathematicians had already in the same year, 1993–4, proved the phenomenon of "mirror duality" within a well-defined class of

[27] It is always possible to *completely* flip the construction, so that *all* Yang–Mills gauge charges are reversed.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

constructions [42, 71, 43]. In turn, this gave rise to a whole "industry" of research [583, 235, 426, 117, 124, 274], and then also the proof that the "mirror duality" – in the general case – is an example of the so-called T-duality [501].

> ***Definition 11.2*** *Two spacetimes, $\mathscr{X}$ and $\mathscr{X}'$, are T-dual if the string model that describes string propagation through the spacetime $\mathscr{X}$ is physically identical to the string model that describes string propagation through the spacetime $\mathscr{X}'$.*

In other words, the spacetimes $\mathscr{X}$ and $\mathscr{X}'$ are T-dual if strings, by propagating through them, do not distinguish between them. In the case when $\mathscr{X} = \mathbb{R}^{1,3} \times \mathscr{Y}$ a $\mathscr{X}' = \mathbb{R}^{1,3} \times \mathscr{Y}'$, T-duality is naturally referring to the $\mathscr{Y}$ and $\mathscr{Y}'$ ("purely" space-like) factors. This relation between the spaces $\mathscr{Y}$ and $\mathscr{Y}'$ is thus indirect, as it is based on an identification of structures between observable quantities, as schematically presented in Figure 11.6. Precisely because of this indirectness are

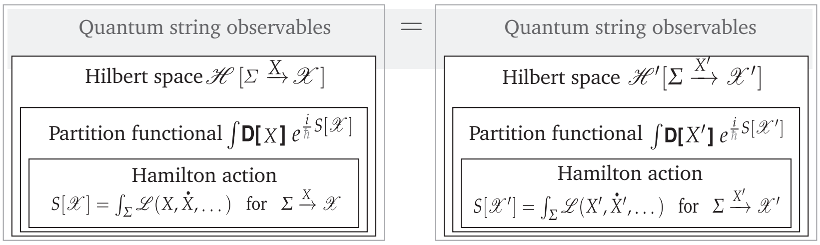


Figure 11.6 A depiction of the indirect relation of "stringy duality" between spacetimes $\mathscr{X}$ and $\mathscr{X}'$.

the so-obtained relations very unexpected, so that the construction and exploration of relations between string models may be regarded also as a machine for generating mathematically non-trivial conjectures, the final proof of which then significantly advances both mathematics and, reciprocally, also physics [455].

Recently, one such physically motivated and unusual general construction [51] of "mirror dual" Calabi–Yau spaces from 1993 was re-examined, providing it with a mathematically more natural formulation 16 years later, and then also with a rigorous proof [320, 108, 72]. It seems worthwhile to note that this construction of "mirror dual" spaces holds even when not all of the defining conditions of Calabi–Yau spaces are fulfilled, and so points to a much more general phenomenon in algebraic geometry, for which there is no other indication, and the mathematical implications of which are only now being explored.

Notice also that the so-defined "stringy duality" is being identified at the level of quantum observables, which are understood to act upon a Hilbert space, and a partition functional for concrete computations. In particular, it is logically not necessary for a model to have a Lagrangian formulation and the associated geometric interpretation. In fact, models have been constructed for which at the time only the partition function was known, and neither a Lagrangian formulation nor a geometrical interpretation was known or needed; see Ref. [203], for starters, and [26, 503, 335] for related recent references.

11.3.3 Variable geometry and cosmology

The basic idea in the Nordstrøm–Kałuża–Klein compactification was the assumption that the entire spacetime has the structure of a product such as $\mathscr{X} = \mathbb{R}^{1,3} \times \mathscr{Y}$. That is, at every point of the first factor there exists an entire copy of the second factor and *vice versa*; also, all copies of the second factor "along" the first factor are identical, and *vice versa*.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

Of course, this may be generalized so that one factor varies from point to point of the other factor. That is, the compact Calabi–Yau space $\mathscr{Y}$ may depend on the location in the non-compact space $\mathbb{R}^{1,3}$. In the general case, such variations define a structure called fibration in mathematics, and the type of the variation provides a finer classification of such constructions. Two examples of fibration are shown in Figure 11.7. In the case of fibering a complex 3-dimensional compact

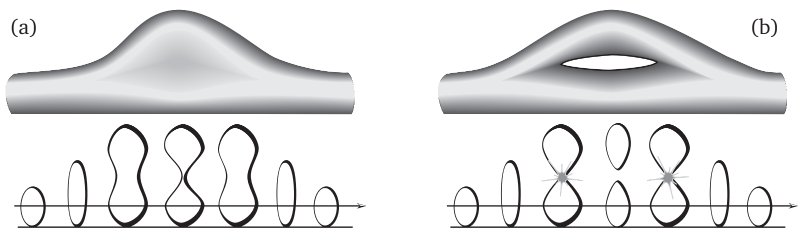


Figure 11.7 Two simple examples of fibration of a loop ($\sim S^1$) along a line: (a) the loop changes its geometry but not its topology, (b) the loop changes both its geometry and its topology ($S^1 \to 2S^1 \to S^1$). In the latter case, there necessarily exist points in the base (horizontal) space where the loop is singular (indicated by the stars).

Calabi–Yau space "along" a $(3+1)$-dimensional spacetime, the situation is of course much more complicated than the simple examples in Figure 11.7 [237, 228]. However, if we suppose that the variations of the compact complex 3-dimensional Calabi–Yau space $\mathscr{Y}$ along the spacetime $\mathbb{R}^{1,3}$ occur subject to certain complex-analytic limitations, it follows that [228]:

1. The compactification space $\mathscr{Y}$ *must* become singular at some spacetime points $\mathrm{x}_* \in \mathbb{R}^{1,3}$, similarly to the situation in Figure 11.7(b).
2. The spacetime locations $\mathrm{x}_* \in \mathbb{R}^{1,3}$ where the Calabi–Yau space $\mathscr{Y}$ becomes singular for a typical $(3+1)$-dimensional observer look like massive objects.
3. The metric tensor in spacetime in the vicinity of these objects has an additional contribution
$$\delta g_{\mu\nu} = (\partial_\mu \phi^a)\, \mathcal{R}_{ab}(\phi)\, (\partial_\nu \phi^b), \tag{11.43}$$
where $\phi^a(x)$ are the scalar fields in spacetime $\mathbb{R}^{1,3}$ (not on the worldsheet Σ), which represent the changes in the complex-analytical structure of the compactifying space $\mathscr{Y}$. The tensor $\mathcal{R}_{ab}$ is the Ricci tensor, computed from the metric tensor $\mathcal{G}_{ab}(\phi)$ given on the target (modular) space in which the scalar field $\phi^a(x)$ takes values.
4. The total number and degree of singularizations may be computed exactly for any concrete model, and is a topological characteristic of the model.
5. With the analytic limitations specified in Refs. [237, 228], these massive objects are lines of cosmic proportions – *cosmic strings* – and affect the distribution of matter (galaxies) in the universe. The gravitational field of filamentary objects in 3-dimensional space decreases as $\sim r^{-1}$, and so dominates in accreting matter from which stars, stellar systems, galaxies and clusters form.
6. Relaxing the analytic limitations [237, 228], the dynamics of these cosmic strings may be analyzed perturbatively, but the total number of interactions (joining and splitting) of these cosmic strings is an exactly computable topological invariant for every model.

Thus, in this rather unexpected way, the details of the (microscopic!) string models also have direct cosmological consequences. The connection between the physics of elementary particles and cosmology is already known even in the popular literature [551], but contemporary research in

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

this area is outside the scope of this book. However, this connection became much more direct in constructing stringy models, whereby the purpose of the following section is to at least offer a sampling from this rich research palette.

11.3.4 Localization of gravity

Courses in electrodynamics show that the discontinuities in the electric and the magnetic fields stem from distributions of electric charges and currents. The coordinate origin, $r = 0$, is in this sense a discontinuity in the radially directed electric field, $\frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r}$. The reason behind the relation between discontinuities in the electromagnetic field and the electric charge and current distributions is of course provided by the Gauss–Ampère laws. The differential equations (5.72a), i.e., (5.78), that represent these laws may be applied both ways:

1. For a given electromagnetic field, we may compute the electric charge and current distribution that produces it.
2. For a given electric charge and current distribution, solve the differential equation and find the produced field.

Regarding gravity and the general theory of relativity, the Einstein equations (9.44) are analogous to the Gauss–Ampère equations. One then expects that discontinuities and other peculiarities in the gravitational field and spacetime curvature may be addressed in an "engineering" fashion, by assembling appropriate distributions of matter [☞ Section 9.3.4].

In typical superstring models, the consistency in dynamics requires that the spacetime through which the superstrings move has 9+1 dimensions. It is then reasonable to ask if it is possible to construct a model where the (9+1)-dimensional spacetime has (3+1)-dimensional isolated subspaces ("defects") of which some may perhaps serve as *our* universe. In such a construction, one must establish:

Q.1 Are there (9+1)-dimensional spacetimes with (3+1)-dimensional "defects"?
Q.2 Are there matter modes that are effectively "trapped" in these "defects"?
Q.3 Do Yang–Mills fields have modes that are effectively "trapped" in these "defects"?
Q.4 Does gravity have precisely one mode that is effectively "trapped" in these "defects"?
Q.5 And, of course, is there at least one (3+1)-dimensional "defect" in which all of the above-cited features occur✌?

Applying Gauss's law within the (3+1)-dimensional spacetime of such defects uses a static subspace of the 3-dimensional space that encloses the source – electric charge or mass – and may be "radially" contracted to the very location of the source. For point-like sources [☞ Section 11.4] in 3-dimensional space, these then are 2-dimensional surfaces, the surface area of which grows with the square of the linear size, whereby the electrostatic and gravitational fields decrease $\sim r^{-2}$ to maintain the constant flux – as we know is the case in Nature.

In turn, if Yang–Mills gauge fields and/or gravity are not trapped in the (3+1)-dimensional defect, the fields will decrease faster: If the field permeates an n-dimensional space, the Gaussian enclosing "wrap" is an $(n-1)$-dimensional sphere, whereupon the Yang–Mills as well as gravitational fields (and forces) decrease following a $\sim r^{1-n}$ law.

— ❦ —

It turns out that the answers to the first three of the above questions are positive, and under very general conditions [280]. Namely, the condition that the compactification space (11.41) is of the Calabi–Yau type (that it admits a metric tensor of which the Ricci tensor is a total derivative), in

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

the worldsheet field theory perspective ("first level" [☞ Section 11.2.4]) becomes a cancellation condition for certain anomalies.

From that same perspective, this same anomaly cancellation condition holds equally for the non-compactified part of the spacetime; for the $\mathbb{R}^{1,3}$ factor (11.41), this anomaly cancellation condition is trivially satisfied. Recall that anomalies are an indication of contradictions in the quantum model – here, in the quantum field theory in the $(1+1)$-dimensional worldsheet spacetime.

The corresponding geometric condition in $(9+1)$-dimensional spacetime field theory ("second level") is that the *entire* spacetime must admit a metric tensor, the Ricci tensor of which is a total derivative. Of course, in the situation (11.41), this is trivially satisfied, since the entire Riemann tensor over the $\mathbb{R}^{1,3}$ factor vanishes. However, this implies that the Wick-rotated $t \to it$ analytic continuation of spacetime may be chosen to be a *non-compact*, complex 5-dimensional Calabi–Yau space.

Existence of large universes with isolated sub-universes

The general property of all Calabi–Yau spaces is that they typically have a large number of nontrivial subspaces, which neither have boundaries themselves nor are the boundary of some other subspace – just like the closed contours A and B in Figure 9.8 on p. 354 and the left-hand illustration in Figure 11.8. The complex 3-dimensional Calabi–Yau spaces – used in compactification (11.41) – have such real 2-, 3- and 4-dimensional subspaces [279]. Complex 5-dimensional Calabi–Yau spaces have such real k-dimensional subspaces with $k = 2, \ldots, 8$. However, such subspaces are not *isolated*, just as the closed paths of type A and B on the surface of a torus are not: each of these contours has continuous deformations/shifts, such as those depicted in the left-hand side of Figure 11.8.

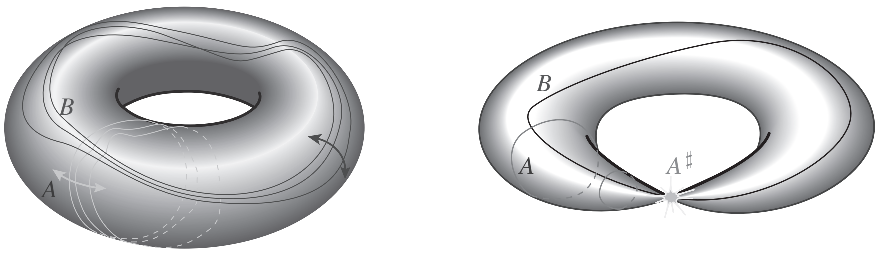


Figure 11.8 The torus surface (left) with two deformable topologically nontrivial closed paths: neither A nor B can be continuously deformed to a point but they can both be deformed into a continuum of "nearby" paths. In the "pinched" torus surface (right), however, the point $A^\sharp$ is isolated [☞ text].

In stark contrast, the surface of the "pinched" torus, on the right-hand side of Figure 11.8 contains the point $A^\sharp$, which is the limiting case of the 1-dimensional subspaces, the closed paths of type A. The point $A^\sharp$ is geometrically and even topologically singled out: any sufficiently small neighborhood of every other point on the surface of the "pinched" torus is of the form of a circle (disc); every neighborhood of the point $A^\sharp$ is of the form of two cones joined at their vertices, and this vertex is a "double point," which is the property that *isolates* this point.

The surface of a torus is in fact a compact complex 1-dimensional Calabi–Yau space, and serves as an intuitive model for higher-dimensional constructions. However, the fact that the double point is also necessarily singular happens only in spaces of one complex dimension. Within complex 3-dimensional Calabi–Yau spaces, such special isolated subspaces are of the form of non-singular real 2-dimensional spheres; they have smooth neighborhoods, but nevertheless cannot be deformed/shifted within the given complex 3-dimensional Calabi–Yau space and so are isolated [279].

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

In general, virtually all complex n-dimensional Calabi–Yau spaces contain special subspaces of complex dimension $\lfloor\frac{n-1}{2}\rfloor$, the integral part of the fraction $\frac{n-1}{2}$. For $n = 5$, these are complex 2-dimensional, i.e., real 4-dimensional subspaces!

Much as a real 2-dimensional surface of a torus serves as an example of a compact complex 1-dimensional Calabi–Yau space, the 2-dimensional cylinder serves as an example of a non-compact complex 1-dimensional Calabi–Yau space. Besides, note that the 2-dimensional sphere has a positive curvature, but that excising two separate points (the 0-dimensional Calabi–Yau space) leaves behind a surface that is a continuous deformation of a cylinder, which is *flat*: there exists a global and single-valued metric tensor for which the Riemann tensor vanishes.

Similarly, every *non-compact* complex n-dimensional Calabi–Yau space may be obtained by excising from a compact complex n-dimensional Fano space[28] a compact complex $(n-1)$-dimensional Calabi–Yau subspace [520, 521]. If the surgery is arranged to also excise part of a special, isolated real 4-dimensional subspace, its remainder is then also non-compact. After a "reverse" $it \to t$ analytical continuation so that one of the four real dimensions is again time-like, these now $(3+1)$-dimensional non-compact subspaces really can serve as isolated examples of $(3+1)$-dimensional spacetime [280].

Conclusion 11.4 *It follows that the analytical continuation of a typical non-compact complex 5-dimensional Calabi–Yau space is a $(9+1)$-dimensional spacetime that contains numerous isolated $(3+1)$-dimensional sub-spacetimes.*

Localization of matter and Yang–Mills gauge interactions

For every example of a complex space Z with a complex algebraic subspace $X \subset Z$, specified as the space of solutions of the system of algebraic equations

$$X \subset Z: \quad X := \{z \in Z,\ \Phi(z) = 0\}, \tag{11.44}$$

there exists a class of restricted functions [☞ [82, 52] and the references therein], defined via the complex n-dimensional generalization of residues,

$$f(x) \ := \ \operatorname*{Res}_{z\in X}\left[\frac{f(z)}{\Phi(z)}\right], \qquad x \in X = \Phi^{-1}(0). \tag{11.45}$$

These functions vanish outside $X \subset Z$, and within X adequately represent fields – both for matter, and also for Yang–Mills gauge fields. It remains of course "merely" to find a concrete non-compact complex 5-dimensional Calabi–Yau space, with a suitable isolated subspace in which (after analytic continuation so that one of the coordinates in X is time-like) the number and type of localized fields can reproduce the contents of the elementary particle physics Standard Model [☞ Table 2.3 on p. 67, as well as Conclusion 2.2 on p. 46].

In the 1990s, string models were routinely constructed containing various p-branes on which, by their very definition, end-points of open strings are trapped on the given p-brane. This then guarantees localized degrees of freedom amongst which it seems realistic to seek the particle content of the Standard Model, including the Yang–Mills gauge fields.

However, owing to significant differences between Yang–Mills gauge fields and the gravitational field [☞ Section 9.2] and since the graviton is inherently realized in string theory by closed strings [☞ discussion that leads to (11.7)], it is not clear that the analysis in Refs. [82, 52] may be adapted so as to be applied to gravity.[29] Thus, the proposal wherein the additional six spatial

[28] after the Italian mathematician, Gino Fano, spaces with a positive curvature are called Fano spaces.

[29] The works of Keiichi Akama and other researchers in Japan and the Soviet Union 1967–82 [☞ [9] and the references therein], where it is shown that an effective general relativity and gravity may be *induced* in $(5+1)$-dimensional models with $(3+1)$-dimensional vortices, until recently were not known outside Japan and the former Soviet Union. However, these models are not renormalizable, and cannot be part of a fundamental theoretical system.

dimensions are not compactified and unobservably small, but it is our (3+1)-dimensional spacetime that is an isolated part – "defect" – within the (9+1)-dimensional spacetime, could not be taken seriously. Compactification *à la* Nordstrøm–Kałuża–Klein remained the only known logical possibility for constructing realistic string (and so also M- and F-theory extended) models, *almost* to the end of the twentieth century.

Localization of gravity

In 1999, Lisa Randall and Raman Sundrum discovered a relatively simple situation in which gravity *is localized* at (a part of) the boundary of a space [450], which "opened vistas and paved avenues" for constructing alternatives to compactification models.

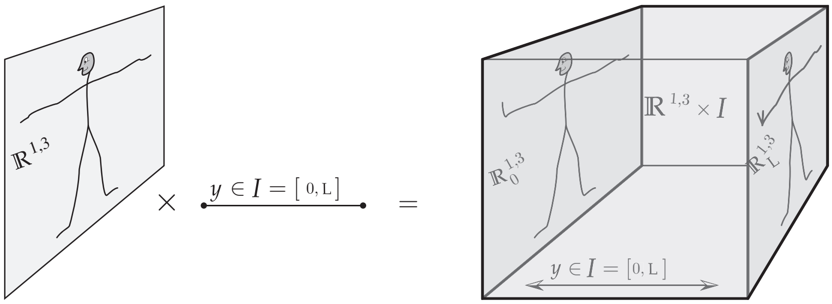


Figure 11.9 The Randall–Sundrum cosmology toy-model.

Randall and Sundrum studied the toy-model wherein spacetime is 5-dimensional geometry reminiscent of a capacitor: the 5-dimensional spacetime is of the form $\mathbb{R}^{1,3} \times I$, where $I = [0, L]$ is the closed interval, i.e., the interval together with its boundary points, as in Figure 11.9. From the definition of the coordinate y, it follows that it can only have non-negative values, and the metric tensor in the 5-dimensional spacetime $\mathbb{R}^{1,3} \times [0, L]$ must depend on $|y|$. However, since the Einstein tensor – the left-hand side of the Einstein equations (9.44) – is a differential expression of *second* order in spacetime derivatives of the metric tensor components, it follows that the Riemann, Ricci and Einstein tensors, as well as the scalar curvature, must include terms proportional to the Dirac δ-function, $\delta(y)$.

Concretely, Randall and Sundrum define

$$\mathrm{d}s^2 = -e^{-2k|y|}\eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu + \mathrm{d}y^2, \quad \text{i.e.,} \quad [\boldsymbol{g}(\mathrm{x}, y)] = \begin{bmatrix} -e^{-2k|y|} & 0 & 0 & 0 & 0 \\ 0 & e^{-2k|y|} & 0 & 0 & 0 \\ 0 & 0 & e^{-2k|y|} & 0 & 0 \\ 0 & 0 & 0 & e^{-2k|y|} & 0 \\ 0 & 0 & 0 & & 1 \end{bmatrix}, \tag{11.46}$$

whereby the Ricci tensor and the scalar curvature are

$$[\boldsymbol{R}] = \left[\begin{array}{c|c} -\eta_{\mu\nu}\, e^{-2k|y|} f(y) & 0 \\ \hline 0 & g(y) \end{array}\right], \qquad \begin{cases} f(y) = 2k\big[\delta(y) - 2k\ \mathrm{sig}^2(y)\big], \\ g(y) = 4k\big[2\,\delta(y) - k\ \mathrm{sig}^2(y)\big], \end{cases} \tag{11.47a}$$

$$R = 16\,k\,\delta(y) - 20\,k^2\ \mathrm{sig}^2(y) = \begin{cases} 16\,k\,\delta(y) & y = 0, \\ -20\,k^2 & y \neq 0. \end{cases} \tag{11.47b}$$

Here

$$\mathrm{sig}(y) := \begin{cases} -1, & y < 0, \\ 0, & y = 0, \\ +1, & y > 0, \end{cases} \qquad \text{so} \qquad \mathrm{sig}^2(y) := \begin{cases} +1, & y \neq 0, \\ 0, & y = 0, \end{cases} \tag{11.48}$$

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

and the results hold in the vicinity of $y=0$, as if $L\to\infty$. The exact result is more complicated and of course must include terms with $\delta(y-L)$ and $\mathrm{sig}^2(y-L)$ because of the analogous effect of the $y=L$ boundary.

The Einstein equations then dictate that $T_{\mu\nu}$ must contain terms proportional to $\delta(y)$, $\delta(y-L)$, $\mathrm{sig}^2(y)$ and $\mathrm{sig}^2(y-L)$. This implies that the maintenance of such a geometry requires the existence of matter that is localized at the $y=0$ and the $y=L$ boundaries of this 5-dimensional "universe," as well as matter that permeates this universe along the fifth, y-coordinate. However, it is more important that the differential equations for the metric tensor components, after separation of variables and a suitable substitution $z=z(y)$, include the differential equation [450],

$$\Big[-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}z^2}+\widetilde{V}_\pm(z)\Big]\hat{\psi}(z)=0,\quad \widetilde{V}_\pm(z)=\frac{15\,k^2}{8(k\,|z|+1)^2}\pm\tfrac{3}{2}\,k\,\delta(z)-\frac{m^2c^2}{\hbar^2},\quad z,y\geqslant 0,\qquad(11.49)$$

with the upper sign at the position $y=L$ and the lower sign at $y=0$. The appearance of the Dirac δ-function in the otherwise rather mildly peaking "potential" reminds us of the familiar system from non-relativistic quantum mechanics. This implies that the case $\widetilde{V}_-(z)$ – i.e., at the $y=0$ copy of the $\mathbb{R}^{1,3}$-like "capacitor" plate in Figure 11.9 on p. 431 – has solutions with the following properties:

1. There exists a single, square-normalizable mode with negative energy, localized at $y=0$, and its amplitude decays exponentially with y.
2. There is a continuum of modes:
   (a) with continuous mass/energy $m^2\in[0,+\infty)$,
   (b) the envelopes (amplitudes) of which are very small near $z=0$ because of the $\frac{15\,k^2}{8(k|z|+1)^2}$ barrier,
   (c) which asymptotically (for $|z|\to\infty$) approach plane waves.
3. Owing to the property 2(b), the interference between the unique localized ("bound-state") mode and the continuum of modes is suppressed.

Randall and Sundrum then showed [450] that the unique localized mode in the metric tensor effectively serves as the metric tensor in the $\mathbb{R}^{1,3}_{y=0}$-boundary of their 5-dimensional model, and leads to a usual formulation of the general theory of relativity in this part of the boundary, as well as to the familiar Newton/Kepler gravitational potential $\sim r^{-1}$. The continuum of modes produces a correction of Newton's law of gravity:

$$V(r)=G_N\frac{M_1\,M_2}{r}\Big(1+\frac{1}{(kr)^2}\Big),\qquad(11.50)$$

where the parameter k is a measure of the curvature (not the size!) of the 5-dimensional spacetime along the fifth coordinate.

Conclusion 11.5 *Comparing equations (11.50) and (11.47b) shows that the correction to Newton's law of gravity is suppressed by the* ***curvature*** *of the "big" spacetime along the fifth dimension, and not the size of this fifth dimension. This result is* ***qualitatively*** *different from similar results in compactification models: There, all corrections that stem from the existence of compact dimensions are always suppressed by the* ***volume*** *of the (small) compact space.*

The fifth coordinate in the Randall–Sundrum model may well be even infinitely large(!); its existence nevertheless changes the effective $(3+1)$-dimensionality of physics in the $y=0$ boundary 3-brane no more than as specified in equation (11.50).

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

This important result started a "minor industry" of elaborations of this and similar ideas, whereby the string theoretical system again forayed into cosmology with these *brane geometry* models. In the 1990s, many details of the interactions between various p-branes and other more-or-less exotic objects that appear in the string theoretical system were worked out. Now that we know of the Randall–Sundrum mechanism for localizing gravity, it is worth exploring the possibility that some of these p-branes – as well as other $(9+1)$-dimensional spacetimes – are of cosmic proportions, and that we happen to live on one of these 3-branes, which provides the basic conceptual idea of the cosmology of so-called brane-worlds [☞ [352], for a recent review].

— ❦ —

Amongst such models an interesting possibility emerges that again links the microscopic and the macroscopic physics in unusual ways: Namely, we know that no supersymmetric partner particle of any of the known particles has ever been found, so – if the fundamental theory of Nature is supersymmetric at all, supersymmetry must be broken, and direct experimental evidence for this (e.g., a Goldstone fermion) is also lacking. On the other hand, the discovery that our universe is expanding in an accelerated fashion implies that the corresponding geometry (in large, cosmic proportions) is the de Sitter geometry. However, it is known [189, 562, 560, 76] that the de Sitter geometry does not admit supersymmetry. Nevertheless, it is possible to construct superstring models containing a $(3+1)$-dimensional sub-spacetime, i.e., 3-brane [53] [☞ also [485, 303, 302] and [16, 120] for recent works]:

1. with the de Sitter geometry,
2. with localized gravity,
3. with an exponential relation between the Planck mass and the mass of $W^{\pm}$- and Z^{0}-bosons,
4. where the geometry is induced by the presence of a modular field,[30]
5. with the cosmological constant related to the supersymmetry breaking [54].

It follows that it is possible to break supersymmetry by means of the spacetime geometry, which in turn is produced by the interaction of gravity with modular fields that are unique to stringy models.

However, this is but one of many possibilities; one "merely" ought to find the model in which on some of its 3-branes with localized gravity there exist enough localized matter and Yang–Mills gauge fields for the Standard Model [☞ Table 2.3 on p. 67, and Conclusion 2.2 on p. 46]. The Reader interested in this class of brane-world models is directed to the rich literature, starting for example with Ref. [452] for relations with strings, and Ref. [101] for F-theory extensions.

Exospace

Besides the two general mechanisms discussed so far,

compactified worlds the Norstrøm–Kałuża–Klein compactification, both the constant (11.41), and the variable kind as discussed in Section 11.3.3,

brane-worlds the Randall–Sundrum mechanism of localizing gravity to some of the sub-spacetimes of a big, $(9+1)$-dimensional spacetime,

there is however also a *third* possibility.

Namely, physics without strings is based on describing the motion of point-like particles ("material points") and their extension, "point-local" fields: Although a field by definition extends and

[30] This is literally a stringy "signature." The particular modular field involved here is not single-valued: rotations in a plane within the "extra" dimensions induce a so-called Möbius, or $SL(2;\mathbb{Z})$ transformation in the field. This occurs in no non-stringy theory/model, and *every* string model contains this particular modular field.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

permeates the entire space, functions that are used to describe fields are fundamentally local quantities. For example, the gauge potential $\mathbb{A}_\mu(x)$ depends on the coordinates of a *single point* in spacetime, and the differential equations that mathematically represent the laws about such fields are local: the fluctuation of the field at any one point in spacetime causes – via the local differential equation of motion – the propagation of the fluctuation from one spacetime point to the infinitesimally neighboring points.

The physics of strings is not local in the same sense. From the worldsheet perspective, the field theory in $(1+1)$-dimensional worldsheet spacetime is not local in the same sense as it is in the $(3+1)$-dimensional spacetime. Students who have successfully passed a course in electrodynamics must know that the Green functions for the wave operator (the d'Alembertian) in $(n+1)$-dimensional spacetime *grow* with the distance when $n < 2$. Thus, scalar fields in $(1+1)$-dimensional spacetime correlate between arbitrarily distant fluctuations, and so are fundamentally global and non-local fields.

On the other hand, from the spacetime perspective in which the strings propagate, it is clear that strings are not local objects, but exist simultaneously (however this to be understood) in a continuum of space-like separated points within the spacetime – this is a property of all extended objects, including also all p-branes with $p > 0$. Besides, the following facts also hold about string interactions:

- String interaction *is* local in the spacetime in which the strings propagate: From any observer's reference system, the joining of two strings into one and the splitting of a string into two happens at one spacetime point.
- String interaction is *not* local in the configuration space of strings; if it were, two strings would be joining into one (and one splitting into two) in a single point in the configuration space – which is a particular configuration of the entire string.
- String interaction is *not* local in the string worldsheet spacetime; moreover, a string interaction represents a "cosmological" fusion of two such $(1+1)$-dimensional spacetimes into one, or the splitting of one into two.

Of course, these are merely picturesque indications that the motion and interactions of strings (in fact, of all p-branes for $p > 0$) differ essentially from those of point-particles (0-branes).

It turns out, however, that these differences are crucial in determining through what kinds of spacetimes strings – and more generally, p-branes with $p>0$ – can consistently propagate. It was already known in 1985 that so-called *orbifolds* – spaces with conical singularities of the form $\mathbb{R}^n/D$ where D is the action of some finite group of rotations – pose no problem [138, 137]. A complete and final criterion to answer the question "through how singular a spacetime can strings consistently propagate" is not yet known☝, but it is known that requiring supersymmetry in stringy dynamics permits singularities of rather high degree [278]. This certainly includes both orbifold and canonical singularities [☞ the "Young Persons' Guide" [454]]. The hallmark property of these types of singularities is that they can be smoothed by means of processes called *blow-up*, *deformation* and *small resolution*, which either maintain or can be restricted to maintain all the characteristics essential for superstring dynamics, such as Ricci flatness [279, for starters].

Closely related to singular spaces are so-called *stratified pseudo-manifolds* through which strings also move consistently [24].[31] Such spaces generalize the cases shown in Figure 11.9 on p. 431, where the 5-dimensional space has 4-dimensional "boundary" parts. In general, stratified pseudo-manifolds are connected unions of several parts organized by dimension so that:

[31] The Authors of Ref. [24] have not emphasized this fact explicitly, but their Figure 19 explicitly depicts a complex 3-dimensional pseudo-manifold with a complex 1-dimensional connected additional part.

1. separately taken, every part is a space of constant dimension,
2. subspaces of the same dimension form a *stratum*,
3. there may exist more than one stratum, i.e., parts of more than one dimension.

Instead of detailed definitions, may it suffice here to consider the two examples in Figure 11.10. The left-hand example is a surface (a), defined by the equation $z = (x/y)^2$, with a self-intersection along the non-negative part of the z-axis, where the surface evidently has two-fold defined tangent vectors and has "unusual" (exotic) neighborhoods. However, this surface may be decomposed into:

0. the (excised) 0-dimensional stratum: the coordinate origin O,
1. the (excised) 1-dimensional stratum: positive z-axis denoted z_+,
2. the (remaining) 2-dimensional stratum: two surfaces, A and B.

In this example, *every* point x_* of every stratum has arbitrarily near points that belong to a higher-dimensional stratum; in general, this need not be true.

The right-hand side of Figure 11.10 shows a more unusual but also more general example (b), which may be decomposed into:

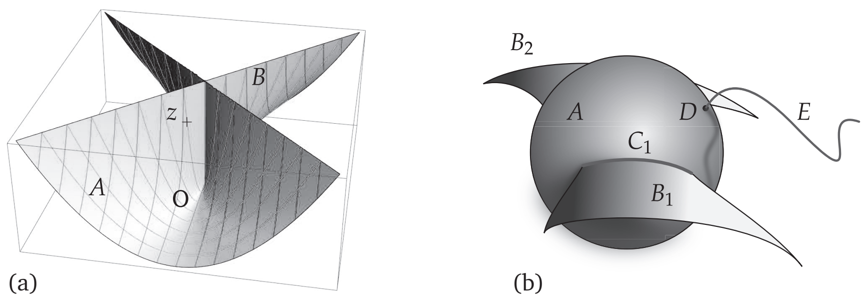


Figure 11.10 Two stratified pseudo-manifolds.

0. the (excised) 0-dimensional stratum: the point D,
1. the (excised) 1-dimensional stratum: the "seams" C_1 and C_2 (on the back side) and the "tail" E,
2. the (remaining) 2-dimensional stratum: two "wings" B_1 and B_2 and the surface of the sphere A with three holes (two line-like cuts where the "seams" C_1 and C_2 were excised, and one point-like hole where the point D was excised).

Unlike example (a), not every point of the 1-dimensional stratum has points in the higher-dimensional stratum (the surface of the sphere A and of the wings B_1 and B_2 and without the "seams" C_1 and C_2) that are arbitrarily close to it. To wit, *each* point in the "seams" C_i is arbitrarily near *some* points in both wings, B_1 and B_2, and in the sphere A. On the other hand, most points in the "tail" E are nowhere near any point in the 2-dimensional stratum. One says that the "tail" E is *outside* the 2-dimensional stratum. Extending the nomenclature of Ref. [24], let the prefix *exo-* denote the *exotic* parts that are mostly *outside* the highest(-dimensional) stratum. As the highest dimension in this example is two (2), the only exo-space (here, exo-line) is the 1-dimensional "tail" E.

Also, the complete pseudo-manifold (b) in Figure 11.10 does not have well (unambiguously) defined derivatives at the "seams" C_1 and C_2 as well as at the joining point D. This is also true of the example (a), along the non-negative semi-axis $z \geqslant 0$.

Finally, the spacetime (so-called *brane geometry*) in a string model can easily be some such stratified pseudo-manifold $\mathscr{Z}$ that contains $(3{+}1)$-dimensional exo-space $\mathscr{X} \subset \mathscr{Z}$, i.e., that $\mathscr{Z}$ admits a metric tensor[32] that gives a 4-dimensional exo-space $\mathscr{X}$ the Lorentzian signature, $(1,3)$. For example, $\mathscr{Z}$ could be caricatured by the right-hand example in Figure 11.10, where $\mathscr{X}$ could be represented by the "tail" E and which could of course easily be of cosmic proportions. To an observer who is in a part of $\mathscr{X}$ sufficiently far from the region where $\mathscr{X}$ joins with the "rest" of the complete spacetime $\mathscr{Z}$, the spacetime evidently looks 3+1 dimensional. However, an observer who is sufficiently close to the connecting region between $\mathscr{X}$ and the "rest of $\mathscr{Z}$" will of course be able to experimentally verify that by passing from $\mathscr{X}$ into the "rest of $\mathscr{Z}$" the number of spacetime dimensions changes. In such a model, all (gravitational and Yang–Mills gauge) fields must be localized, simply as they have nowhere else to propagate in the neighborhood of almost all points in $\mathscr{X}$. Then, the Gaussian surface that encloses point-like sources (charges and/or masses) may almost everywhere in $\mathscr{X}$ be chosen to be a sphere of surface $4\pi r^2$ at a distance r from the source, whereby the Coulomb and the Newton/Kepler forces indeed decrease as $\sim r^{-2}$. The exploration of such models is in its infancy☝, but it is clear that in many models the p-branes may well be *exo*-branes, just like in Figure 19 of Ref. [24] or in the right-hand example in Figure 11.10. For mathematical details about stratified pseudo-manifolds, the Reader is directed to the literature, e.g., starting with the book [313].

11.3.5 Exercises for Section 11.3

✎ **11.3.1** *For the metric tensor in* $(2{+}1)$*-dimensional spacetime*

$$ds^2 = -e^{-2k|y|}c^2dt^2 + e^{-2k|y|}dx^2 + dy^2, \quad \text{i.e.,} \quad [g(x,y)] = \begin{bmatrix} -e^{-2k|y|} & 0 & 0 \\ 0 & e^{-2k|y|} & 0 \\ 0 & & 1 \end{bmatrix}, \tag{11.51}$$

compute (for $m,n,p,r = 0,1,2$*):*

1. *the Christoffel symbol* Γ^p_{mn}*;*
2. *the Riemann tensor* $R_{mnp}{}^r$*;*
3. *the Ricci tensor* R_{mn}*;*
4. *the scalar curvature* R_{mn}*;*
5. *the Einstein tensor* $G_{mn} := R_{mn} - \frac{1}{2}g_{mn}R$.

✎ **11.3.2** *Show by direct computation that the change* $e^{-2k|y|} \to e^{-2ky}$ *in the metric tensor definition in the previous exercise "erases" all matter localized in the spacetime plane* $y = 0$ *and that the results for* Γ^p_{mn}, $R_{mnp}{}^r$, R_{mn}, R *and* G_{mn} *agree with the formal substitution* $\mathrm{sig}^2(y) \to 1$ *in the results of the previous exercise.*

11.4 Duality and dual worldviews

Section 5.2.2 showed that there exists a symmetry between the electric field, charge and current on one hand, and the magnetic field, charge and current on the other [☞ Conclusion 5.4 on p. 185]. The fact that – as best as known – there are no magnetic monopoles in Nature, i.e., monopole magnetic charges and currents means that the duality rotation (5.85), $\varpi_{EM}(\vartheta)$, may be applied simultaneously both to electromagnetic fields and to electromagnetic sources so that this basis is – everywhere in the universe "simultaneously"! – "turned" orthogonal to $(\rho_m/c, \vec{\jmath}_m/c^2)$ so that only $(c\rho_e, \vec{\jmath}_e)$ remain. However, the *same* system can be equally well described in the "basis"

[32] The metric tensor may be defined separately in each part of constant dimension, and then require that these tensors coincide in places where the various parts touch.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

that is "turned" orthogonal to $(c\rho_e, \vec{\jmath}_e)$ so that only $(\rho_m/c, \vec{\jmath}_m/c^2)$ remain, as was summarized in Conclusion 5.4 on p. 185. This then provides two *dual* descriptions of Nature.

Exploration of the string theoretical system discovered many other such *dual* relations between many (sometimes very) different models. In this sense, one says:

Definition 11.3 *Two given descriptions of Nature, O_1 and O_2, are* ***dual*** *if they are* ***physically indistinguishable****, i.e., if the collection of all physical observables (together with all relations between them) in the description O_1 is isomorphic to the corresponding collection in the description O_2.*

(There may be more than two dual descriptions, so one talks of ***triality****, but in practice one never talks of* ***quadrality****,* ***quintality****, or any other* ***n-ality****.)*

Given the complexity of the relations between the physical system, the mathematical model, the solutions to that model and measurable (verifiable) results [☞ Figure A.2 on p. 457], the limitations on the mathematical model itself are evidently very indirect and roundabout. It should then be clear that the mental caricatures and images that one uses in formulating the model are merely a crutch in the construction of the model, and not "the one and true" image of "reality" – recall the Copernican legacy [☞ Section 1.1.1]. Thus, as long as two or more models (even if based on different images) equally well describe Nature, we are free to choose which of these two (or more) images to tentatively identify with Nature [☞ Section 11.3.2, and especially Figure 11.6 on p. 426]. In doing so, we must stick to the facts [☞ Example 1.1 on p. 11]:

1. The chosen image is but one of the a-priori equally "real" formulation images.
2. The choice of the formulation image is, without an experimentally verifiable difference, subjective and tentative.
3. Measurable results (and not the formulation image) of a model make up its goal, and so also its point.

The last two decades made it ever clearer that the string theoretical system integrally contains multiple differing formulation images in describing the same "thing" – which lends support to conceiving of a new (unexplored and undeveloped☝) kind of "symmetry," acting via corresponding transformations between different formulation images, and which are *not* the usual gauge symmetries [126, 77, 470].

11.4.1 T-duality

The first examples of duality were discovered more-or-less accidentally, whereby this research reminds us of experimental physics where the first task is to find as many examples as possible so as to perceive the common properties, so as to then seek the basic principles of this phenomenon. Although there exist many seemingly different examples of T-duality, it turns out that this is one of the simpler classes of duality, and that other types of duality are progressively more and more unusual – from the vantage point of the well-known general properties in previously known physical models. This of course merely points to the fact that the (M- and F-theory extended) (super)string theoretical system has radically new, unknown and unexpected properties.

$R \to 1/R$ duality

The first signal of this multiplicity is the so-called $R \to 1/R$ duality, which is described in detail in textbooks [434, 594, 46]. This is an essential consequence of the fact that strings are not point-like and that the relation between strings and the spacetime through which they move is very different from the analogous relation for point-like particles.

Let the $\hat{\imath}$th spacetime coordinate (for some fixed $\hat{\imath} > 0$) be periodic and closed into a circle of circumference $2\pi R$. To a string that propagates through such a spacetime, the coordinate field in this periodic direction must satisfy the same periodicity condition:

$$X^{\hat{\imath}}(\tau,\sigma) \simeq X^{\hat{\imath}}(\tau,\sigma) + 2\pi w\, R, \qquad w \in \mathbb{Z}, \tag{11.52}$$

where w counts how many times the string is wrapped around the circle of the periodic coordinate. This periodicity changes neither the Hamilton action nor the Lagrangian density, but imposes the linear momentum quantization in the direction of the periodic (compact) coordinate:

$$p_{\hat{\imath}\, n} = \frac{n\hbar}{R}, \qquad n \in \mathbb{Z}, \quad \text{fixed } \hat{\imath}. \tag{11.53}$$

The proper generalization of the relation (3.36) gives the square of the Lorentz-invariant mass of that string [434, 594] as

$$m^2c^2 = \frac{n^2\hbar^2}{R^2} + \frac{w^2R^2}{\alpha'^2\hbar^2c^4} + \frac{2}{\alpha' c^2}\big(N_L + N_R - 2\big), \tag{11.54}$$

where N_L and N_R are the total excitation numbers for the left-moving and the right-moving oscillators (11.38), respectively, counting only oscillators that are transversal to the worldsheet and which satisfy the condition

$$nw + N_R - N_L = 0. \tag{11.55}$$

Manifestly, the relations (11.54)–(11.55) remain unchanged under the exchange

$$n \leftrightarrow w, \qquad \text{and} \qquad R \leftrightarrow \frac{\alpha'\hbar^2c^2}{R} = \frac{{\ell_S}^2}{R}, \tag{11.56}$$

which implies that string models do not distinguish between spacetimes in which a periodic coordinate describes a circle of circumference $2\pi R$ and those with a circle of circumference $2\pi({\ell_S}^2/R)$. For a complete proof of this equivalence according to the diagram in Figure 11.6 on p. 426, see Chapter 8 in the textbook [434, Vol. 1].

Manifestly, if $R < \ell_S$, then $({\ell_S}^2/R) > \ell_S$. Thus, a spacetime with a compact dimension that is smaller than the string characteristic size, $\ell_S \sim 10^{-35}$ m, is equivalent to a spacetime where the compact dimension is reciprocally larger than this characteristic length. In this sense, compact dimensions cannot be "too small" to be "seen" by strings – which is exactly the opposite behavior from that in "pointillist" models, for which $\ell_P \sim 10^{-35}$ m is the minimal discernible distance [☞ Section 1.3].

This duality between "big" and reciprocally "small" dimensions is called the T-duality, in that the target space in which the coordinates are all periodic has the geometry of a torus, $T^n := S^1 \times \cdots \times S^1$ (with n factors).

Mirror duality, again

Mirror duality, discussed in Section 11.3.2, was first discovered as a relation between two Calabi–Yau compactification models that use two very concrete constructions [236]: the first Calabi–Yau manifold, X, is the space of solutions to the algebraic equation

$$\sum_{i=1}^{5} {z_i}^5 = 0, \qquad z_i \in \mathbb{C}, \quad (z_1,\dots,z_5) \simeq (\lambda z_1,\dots,\lambda z_5), \quad 0 \neq \lambda \in \mathbb{C}. \tag{11.57}$$

The other Calabi–Yau manifold is obtained as the quotient space $Y = X/(\mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_5)$, where each factor $\mathbb{Z}_5$ denotes an independent symmetry of the fifth order, such as the transformation

$$(z_1,z_2,\dots,z_5) \simeq (\omega^1 z_1, \omega^2 z_2, \dots, \omega^5 z_5), \qquad \omega := e^{2\pi i/5}, \;\; \omega^5 = 1. \tag{11.58}$$

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

It may be computed [236, 279] that $h^{1,1}(X) = 1 = h^{2,1}(Y)$ and $h^{2,1}(X) = 101 = h^{1,1}(Y)$, which is an indication of the "mirror duality" $X \leftrightarrow Y$. The complete proof [236] is much more detailed and requires showing that compactifications on the manifolds X and Y are physically equivalent, in the sense of the diagram in Figure 11.6 on p. 426. This construction of an explicit pair of mirror-dual Calabi–Yau manifolds was soon generalized in several different ways, and Andy Strominger, Shing-Tung Yau and Eric Zaslow had by 1996 proven that all such pairs are special cases of T-duality [501]. Complementary to this research, done from the perspective of the spacetime through which the string moves, models in $(1{+}1)$-dimensional worldsheet spacetime were also soon constructed that completely reproduce the mirror duality [188, 371, 192].

— ❦ —

In all these examples, many of the geometric and topological properties of the space X and its T-dual space X^T are different, but $\dim(X^T) = \dim(X)$. Even this need not be the case in other types of duality.

11.4.2 Gauss's law and its consequences

The insight that the Einstein equations are the analogue of Gauss's law for the gauge symmetry of general coordinate reparametrizations motivates a re-examination of Gauss's law for the various possible cases.

Electric and magnetic sources in 4-dimensional spacetime

Electric charge is – in principle – measured by means of measuring the total flux of the electric field through a closed (Gaussian) surface that encloses the given charge distribution. From the practical application of this (Gauss's) law in electrostatics, we know that for the electric field the smallest possible electric charge is point-like, i.e., 0-dimensional. Of course, a collection of point-like charges may well form a linear, surface or volume distribution of charges, as most often observed in Nature. However, the point here is that nothing in the structure of electrodynamics obstructs an electric charge from being as little as 0-dimensional.

Analogously, the same thought-experiment/measuring may also be set up for a magnetic field, and it follows that the smallest possible magnetic charge is also point-like – as discussed in Section 5.2.3. The fact that point-like magnetic charges (magnetic monopoles) are not observed in Nature is then a puzzling property of *our* particular Nature, even if we take into account the relativity of what is called "electric" and what "magnetic" [☞ Conclusion 5.4 on p. 185].

In preparation to re-examine this fact in the more general case, consider in a bit more detail how the application of Gauss's law produces the minimal dimension (= 0) for electric and magnetic charges in $(3{+}1)$-dimensional spacetime.

Using the 4-vector notation (5.73), the components of the electric field are identified as $E_i = F_{0i}$. The flux of the electric field is maximal when measured (integrated) over a surface of which the tangent plane at every point is orthogonal to the direction of the electric field. (As the electric field is a 3-vector, in 3-dimensional space the orthogonal subspace is of course a 2-dimensional surface.) The Gaussian integral over the closed surface (2-dimensional sphere) is then written

$$\Phi_E := \oint_{S_G} \mathrm{d}^2\vec{\sigma}\cdot\vec{E} = \oint_{S_G} \mathrm{d}^2\sigma^i\, E_i = \oint_{S_G} \mathrm{d}^2\sigma^i\, F_{0i} = \oint_{S_G} \mathrm{d}x^\mu \mathrm{d}x^\nu\, \varepsilon_{\mu\nu}{}^{0i}\, F_{0i}. \tag{11.59}$$

Owing to the defining properties of the Levi-Civita symbol, $\varepsilon_{\mu\nu}{}^{\rho\sigma} := \varepsilon_{\mu\nu\kappa\lambda}\eta^{\kappa\rho}\eta^{\lambda\sigma}$, we know that $\mu,\nu \neq 0, i$, i.e., that $\mathrm{d}x^\mu$ and $\mathrm{d}x^\nu$ are differentials of coordinates that are, in every point of the Gaussian surface S_G, orthogonal to the direction of the electric field E_i, as well as to the direction of time – i.e., they are constant in time. As the electric field $\vec{E}$ is directed *from* the source (positive electric charge) or *towards* the sink (negative electric charge), the Gaussian surface S_G can then be

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

shrunk "radially" – along the ith coordinate in equation (11.59) – to the source/sink of the $\vec{E}$-field, electric charge itself. The dimension of this source/sink then must be equal to

$$\dim(\rho_e) \geqslant \dim(\text{space}) - \dim(S_G) - \dim(\text{radius}) = 3 - 2 - 1 = 0. \tag{11.60}$$

That is, electric charges may well be as little as 0-dimensional (point-like).

The components of the magnetic field are identified as $B^i := \frac{1}{2}\varepsilon^{ik}F_{jk} = \frac{1}{2}\varepsilon^{0ik}F_{jk}$, i.e., $F_{jk} = B^i\varepsilon_{0ik}$, so unlike equation (11.59), for the magnetic flux we have

$$\begin{aligned}\Phi_M &:= \oint_{S'_G} \mathrm{d}^2\sigma_i\, B^i = \oint_{S'_G} \mathrm{d}x^i \mathrm{d}x^j\, B^k\, \varepsilon_{0ijk} = \oint_{S'_G} \mathrm{d}x^j \mathrm{d}x^k\, F_{jk} \\ &= \oint_{S'_G} \mathrm{d}x^\mu \mathrm{d}x^\nu\, \varepsilon_{\mu\nu 0i}\big(\tfrac{1}{2}\varepsilon^{0ijk}\, F_{jk}\big) \overset{(5.85)}{=} \oint_{S'_G} \mathrm{d}x^\mu \mathrm{d}x^\nu\, \varepsilon_{\mu\nu 0i}\big(*F^{0i}\big).\end{aligned} \tag{11.61}$$

Again, the use of the Levi-Civita symbol ε^{0ijk} guarantees that $\mathrm{d}x^i$, $\mathrm{d}x^j$ and B^k are mutually orthogonal. The Gaussian surface S'_G again may be shrunk "radially" to the magnetic charge itself, the dimension of which then must be equal to

$$\dim(\rho_m) \geqslant \dim(\text{space}) - \dim(S'_G) - \dim(\text{radius}) = 3 - 2 - 1 = 0. \tag{11.62}$$

Thus, in (3+1)-dimensional spacetime the sources (or sinks) of both electric and magnetic fields may be as little as 0-dimensional (point-like). Notice the similarity between equations (11.59) and (11.61) owing to the fact that both $F_{\mu\nu}$ and its dual, $*F^{\mu\nu}$, are rank-2 tensors.

Sources for gauge fields in *n*-dimensional spacetime

The above analysis probably seems unnecessarily complicated for such an "obvious" result. However, in more than (3+1)-dimensional spacetime, the results are less obvious.

In n-dimensional spacetime, the electric flux is

$$\Phi_E := \oint_{S_G} \mathrm{d}x^{\mu_1} \cdots \mathrm{d}x^{\mu_{n-2}}\, \varepsilon_{\mu_1\cdots\mu_n}\eta^{\mu_{n-1}0}\eta^{\mu_n i}\, F_{0i}, \tag{11.63a}$$

$$\dim(\rho_e) \geqslant (n-1) - (n-2) - 1 = 0: \quad \Rightarrow \ \text{point-like electric charges.} \tag{11.63b}$$

The magnetic flux (writing $(*F)^{\rho_1\cdots\rho_{n-2}} := \frac{1}{2}\varepsilon^{\rho_1\cdots\rho_{n-2}jk}\, F_{jk}$) is

$$\Phi_M := \oint_{S_G} \mathrm{d}x^\mu \mathrm{d}x^\nu\, \tfrac{1}{(n-2)!}\varepsilon_{\mu\nu\rho_1\cdots\rho_{n-2}}\, (*F)^{\rho_1\cdots\rho_{n-2}}, \tag{11.64a}$$

$$\dim(\rho_m) \geqslant (n-1) - (2) - 1 = n-4: \quad \Rightarrow \ (n-4)\text{-dimensional magnetic charges.} \tag{11.64b}$$

Thus, in spacetime of more than $3+1$ dimensions, the magnetic sources/sinks may no longer be point-like. This follows from the fact that in n-dimensional spacetime the components of the magnetic field are identified with the components of the *dual* tensor:

$$E_i := F_{0i}, \qquad \text{but} \qquad B^{i_1\cdots i_{n-3}} := \varepsilon^{0i_1\cdots i_{n-3}jk}F_{jk}, \tag{11.65}$$

so the components of the electric field always form a spatial vector, but the components of the magnetic field form a spatial rank-$(n-3)$ tensor: A rank-r tensor "emanates" radially from the source (or towards a sink) that is therefore $(r-1)$-dimensional.

In string models for the first time there routinely also appear gauge fields for which the gauge potential itself is an antisymmetric rank-r spacetime (Lorentz) tensor, $A_{\mu_1\cdots\mu_r}(\mathrm{x})$. The gauge fields are then defined using the rank-$(r+1)$ tensor:

$$E_{i_1\cdots i_r} := F_{0i_1\cdots i_r}, \qquad\qquad A_{\mu_1\cdots\mu_r} := A_{[\mu_1\cdots\mu_r]}, \tag{11.66}$$

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

$$B^{i_1\cdots i_{n-r-2}} := \varepsilon^{0\,i_1\cdots i_{n-r-2}\,j_1\cdots j_{r+1}} F_{j_1\cdots j_{r+1}}, \qquad F_{\mu_1\mu_2\cdots\mu_{r+1}} := \partial_{[\mu_1} A_{\mu_2\cdots\mu_{r+1}]} \tag{11.67}$$

where the square brackets around the indices denote total antisymmetrization [☞ the lexicon entry, in Appendix B.1]:

$$a_{[\mu}b_{\nu]} := \tfrac{1}{2}\big(a_\mu b_\nu - a_\nu b_\mu\big) = \tfrac{1}{2!(n-2)!}\varepsilon_{\mu\nu\rho_1\cdots\rho_{n-2}}\varepsilon^{\kappa\lambda\rho_1\cdots\rho_{n-2}}a_\kappa b_\lambda, \tag{11.68a}$$

$$a_{[\mu}b_\nu c_{\rho]} := \tfrac{1}{3!}\big(a_{[\mu}b_{\nu]}c_\rho + a_{[\rho}b_{\mu]}c_\nu + a_{[\nu}b_{\rho]}c_\mu\big)$$
$$= \tfrac{1}{3!(n-3)!}\varepsilon_{\mu\nu\rho\sigma_1\cdots\sigma_{n-3}}\varepsilon^{\iota\kappa\lambda\sigma_1\cdots\sigma_{n-2}}a_\iota b_\kappa c_\lambda, \qquad \text{etc.} \tag{11.68b}$$

These imply the generalization of the relations (11.63b) and (11.64b):

Conclusion 11.6 *The minimal dimensions of the electric and magnetic charges in n-dimensional spacetime for gauge interactions with a rank-r gauge potential are*

$$\dim(\rho_e) \geqslant \big((n-1)\big) - \big(n-(r+1)\big) - 1 = r - 1: \quad \textit{electric } (r-1)\textit{-branes;} \tag{11.69a}$$

$$\dim(\rho_m) \geqslant \big((n-1)\big) - \big(r+1\big) - 1 = n - r - 3: \quad \textit{magnetic } (n-r-3)\textit{-branes.} \tag{11.69b}$$

For neither of these to become negative, it follows that

$$0 \leqslant \big[p := (r-1)\big] \leqslant (n-4). \tag{11.70}$$

The upper limit is then always taken to be $(n-4) = 7$, corresponding to $n = 11$ in the M-theory extension of string theory. It is not clear if the F-theory extension could also permit 8-branes, but 7-branes certainly do play a key role in the original definition of F-theory [530].

Comment 11.2 *Note that all branes have a tension that is determined by a relation of the type (11.7), and that the largest branes have $n-4$ spatial dimensions. Such branes can then "trap" $n-4$ dimensions of space, obstructing their expansion during the Big Bang, which provides the possibility of a* ***dynamical*** *explanation of the fact that only 4 dimensions of spacetime have characteristic scales of cosmic proportions, while the remaining $n-4$ spatial dimensions may have a size of the order of the Planck length, ℓ_P – thus supporting the compactification type of spacetime geometry.*

*This image is supported by the example of the vibrations and motion of a closed string (as a 1-brane): The vibrations of the string are not limited by the string tension except in the "radial" direction, which would significantly change the length/circumference of the string itself. The range of motion of a closed string may thus be parametrized by **(1)** a "radial" coordinate that is effectively and naturally "trapped" to be within the order of magnitude of ℓ_P, and **(2)** one or more "transversal" coordinates that are not so limited. In this perspective [☞ Section 11.2.3], spacetime is indeed* ***spanned/generated*** *by the modes of string motion/oscillation, and the range of these oscillations thus describes the geometry of this generated spacetime as effectively compactified in the "radial" direction but flat in the "transversal" directions.*

Thus, the electro-magnetic duality then implies that the observable dynamics of electric $(r-1)$-branes is dual to the dynamics of magnetic $(n-r-3)$-branes – although these are in most cases objects of differing dimensions. For example, for rank-2 antisymmetric tensor gauge potentials, $A_{\mu\nu} = -A_{\nu\mu}$, so that $F_{\mu\nu\rho} := (\partial_{[\mu}A_{\nu\rho]}) = \frac{1}{3}(\partial_\mu A_{\nu\rho} + \partial_\nu A_{\rho\mu} + \partial_\rho A_{\mu\nu})$, we have that $r = 2$, so

$$\dim(\rho_e) \geqslant 1, \qquad \dim(\rho_m) \geqslant (n-5). \tag{11.71}$$

Electric charges are then at least 1-dimensional (linear, filamentary distributions), and magnetic charges are at least $(n-5)$-dimensional – point-like in $(4+1)$-dimensional spacetime, but linear in $(5+1)$-dimensional spacetime, etc.

For all these generalized abelian (commutative) fluxes, one defines

$$\mathbf{d} := \mathrm{d}x^\mu \partial_\mu, \qquad \mathbf{A}_{(r)} := \mathrm{d}x^{\mu_1} \cdots \mathrm{d}x^{\mu_r} A_{\mu_1 \cdots \mu_r}(\mathrm{x}), \tag{11.72a}$$

$$\mathbf{F}_{(r+1)} := \mathbf{d}\mathbf{A}_{(r)} := \mathrm{d}x^{\mu_1} \cdots \mathrm{d}x^{\mu_{r+1}} \big[F_{\mu_1 \cdots \mu_{r+1}}(\mathrm{x}) := \big(\partial_{[\mu_1} A_{\mu_2 \cdots \mu_{r+1}]}(\mathrm{x})\big)\big]. \tag{11.72b}$$

Then, by direct generalization of the Dirac dual charge quantization condition (5.112), one obtains

$$q_e^{(r)} \int_S \mathbf{F}_{(r+1)} = q_e^{(r)} \oint_{\partial S} \mathbf{A}_{(r)} \overset{!}{=} 2\pi\, n_{(r)} \in 2\pi\mathbb{Z}, \qquad \forall S, \quad \dim(S) = (r+1). \tag{11.73a}$$

However, the duality between the magnetic and electric fields in $\mathbf{F}_{(r+1)}$ then implies also

$$q_m^{(n-r-2)} \int_{S'} ({*}\mathbf{F})_{(n-r-1)} = q_m^{(n-r-2)} \oint_{\partial S'} (\widetilde{\mathbf{A}})_{(n-r-2)} \overset{!}{=} 2\pi\, n_{(n-r-2)} \in 2\pi\mathbb{Z}, \tag{11.73b}$$

for every $(n-r-1)$-dimensional subspace $S' \subset \mathscr{X}$ of the spacetime $\mathscr{X}$ in which the magnetic charge is or moves; following Comment 5.6, we define $\widetilde{\mathbf{A}}$ to satisfy ${*}\mathbf{F} = \mathbf{d}\wedge\widetilde{\mathbf{A}}$.

These two quantization relations (11.73) of course produce the same generalization of the result (5.108). The important novelty follows from its application in cases when the total spacetime is a product, $\mathscr{X} = \mathscr{X}' \times \mathscr{Y}$, and the pair of quantization conditions (11.73) may be applied using (iteratively) the dualization (denoted by the symbol $*$) *independently* within either one of the three spaces, $\mathscr{X}$, $\mathscr{X}'$ and/or $\mathscr{Y}$. For example, if $\dim(\mathscr{X}) = 10$, $\dim(\mathscr{X}') = 4$ and $\dim(\mathscr{Y}) = 6$, we have

$$\mathbf{F}_{(r+1)}, \qquad ({*}_{\mathscr{X}}\mathbf{F})_{(10-r-1)}, \qquad ({*}_{\mathscr{X}'}\mathbf{F})_{(4-r-1)}, \qquad ({*}_{\mathscr{Y}}\mathbf{F})_{(6-r-1)} \tag{11.74}$$

at our disposal, as well as the *independent* charge-quantization for all these fields. This property was first noticed in 1996 [500], and four years later it was shown that this condition, within the string theoretical system, implies the quantization of many quantities that are continuous in pointillist theories. The resulting discrete available space of models is thus dubbed (in distinction from continuum) *discretuum* [74].

Comment 11.3 *Interestingly, string models are being applied, fully in the spirit of Section 11.1, also in areas of physics that seemingly have no relation with either (relativistic) elementary particle fundamental physics or cosmology, but where experiments are possible and even rather easily accessible [☞ for example, Refs. [529, 337], for starters].*

Swamp and landscape

In the first decade of the twenty-first century, a trend was noticed in several general properties in the (M- and F-theory extended) (super)string theoretical system, which distinguishes this theoretical system from the pointillist theoretical system.[33] Conceptual differences are mostly in favor of the stringy models, although they are technically (much more) demanding.

Wherever possible, one can compare the "volume" of the space of possible models within the string theoretical system with the corresponding result for the pointillist models. For example, in a model where a scalar field appears, one may inquire how big is the space in which the scalar field takes values. To this end, we need a preferred "volume" measure, which usually follows from

[33] The nomenclature is necessarily imprecise here: The theoretical system of strings necessarily also includes various *irreducible* p-branes (of various origins, of various properties and also for various values of $0 \leqslant p \leqslant 7$ or perhaps $\leqslant 8$), and even 0-branes that are really point-like objects; "theoretical system of strings and things" thus does seem to be a nitpickingly correct name. By contrast, in pointillist models, all spatially extended objects (and charge distributions as well as fields) may be reduced to collections of functions that depend on the coordinates of only one spacetime point.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

the (preferred) choice of the metric tensor on the space of values of this scalar field, and the same choice then also dictates the dynamics in the given model. (Notice that such questions were rarely if ever raised before the advent of string theory.) The "physically preferred choice" is thus obtained by reverse reasoning:

$$\text{field dynamics} \;\to\; \begin{matrix}\text{metric on the}\\ \text{space of field values}\end{matrix} \;\to\; \begin{matrix}\text{volume element for the}\\ \text{space of field values.}\end{matrix} \tag{11.75}$$

Even in the case of a periodic dimension of radius R [☞ first part of Section 11.4.1], strings effectively "reduce" the space of possible radii, $\mathcal{M}(R)$, to the semi-infinite interval $R \in [\ell_S, \infty)$, which is equivalent to the interval $R \in (0, \ell_S]$, and the "volume" of which $\mathrm{Vol}(\mathcal{M}(R)) = |\int_0^{\ell_S} \frac{\mathrm{d}r}{r}| = |\int_{\ell_S}^{\infty} \frac{\mathrm{d}r}{r}|$ diverges but only logarithmically, and only at the limit $R \to 0$, which is dual to the infinitely large radius where periodicity fails to make sense. By contrast, pointillist models have no reason for excluding any part of the whole interval $R \in [0, \infty)$, except that in the limiting case $R \to \infty$ periodicity fails to make sense; however, the limiting case $R \to 0$ now effectively corresponds to a new model with spacetime of one fewer dimensions.

The interval $[0, \infty)$ is thus the parametric space of every pointillist model with one periodic coordinate – together with the limiting case $R = 0$, which in fact represents a radically different model. By contrast, the interval $[\ell_S, \infty)$ is the parametric space for stringy models with one periodic coordinate – and does not include the lower-dimensional "stowaway."

Even better motivated is the case of a so-called *modulus*[34] in Calabi–Yau compactifications, including the so-called dilaton–axion (complex) scalar field, where various dualities help to reduce the otherwise infinite space of choices of values for this field to a space of finite volume. Besides, every concrete compactification has a finite number of parameters, and so – for every compactification model – the total number of choices is described by a space of which the volume in the Weil–Petersson–Zamolodchikov metric [89, and references therein] is finite [522, 348], just as was the case for the much simpler torus compactifications, where $\mathscr{Y}$ in the decomposition (11.41) is a real 6-dimensional torus.

Conclusion 11.7 *From the ever more rigorously verified property that the (M- and F-theory extended) (super)string theoretical system consists of a* ***discretuum*** *(and not a continuum) of models,*[35] *it follows that the string theoretical system is a far better defined theoretical system than any pointillist theoretical system.*

This induced the picturesque vision [531] where the vast majority of models that can be constructed within classical, (in various ways) incompletely quantum and/or quantum but non general-relativistic physics form a *swamp*, from which emerge the models that are completely quantum and general-relativistically consistent, and which form the *landscape*. These latter models, or at least their non-empty subset, one believes are string models.

Of course, for *our* Nature, one believes it to be described by some model within the *landscape*, and the problem is "only" that there are so many models that one does not know where to start looking.

AdS/CFT, i.e., gravity/gauge duality

The idea of duality has, for the first time, clearly been manifested in the example of the particle–wave duality in the creation of standard quantum physics – which belongs to the pointillist

[34] *Moduli* are scalar fields, the expectation values of which parametrize the geometrical characteristics of the compactification Calabi–Yau spaces, such as the choice of the complex structure and of the complexified Kähler class. More generally, the analogous reasoning may apply to all *parameters* in stringy models [☞ the structure (11.35)].

[35] Conceptually, all limitations on the theoretical system that indicate the discreteness of the string theoretical system follow from a combination of quantumness and general relativity of Nature.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

theoretical system. The electro-magnetic duality – although in fact part of the classical theory – has acquired a wide application only through its generalization within stringy models and their M- and F-theory extensions, where it indicates that neither the dimension of mathematical objects used to represent physical objects nor the dimension of spacetime in which these objects move and interact are inviolable and sacrosanct constants.

Within the theoretical system of strings, however, a duality was discovered in 1997 that relates string models with point-particle quantum field theory, the so-called AdS/CFT duality [354, 440]. By now, this duality has been generalized to many other examples, into a general gravity/gauge duality [☞ lecture notes [437], where this phenomenon was related to a loophole in the Weinberg–Witten theorem 6.1 on p. 249]. The general characteristic of the gravity/gauge dual examples is the identity between the physical observables and relations between them, for two models that were obtained as different limiting cases of the same superstring model, where

1. one limiting case represents a *superstring* model with the spacetime geometry that contains an $(n+1)$-dimensional anti de Sitter factor [☞ definition (9.81)],
2. the other limiting case is a *point-particle* supersymmetric gauge theory, the degrees of freedom of which are "trapped" at the $(n-1)+1$-dimensional conformal boundary (with the Minkowski metric) of the anti de Sitter space [☞ expression (9.83) and the related discussion].

Contemporary (at the beginning of the twenty-first century) understanding of the string theoretical system contains both the consistent and the absolutely comprehensive application of both (**1**) the gauge principle and (**2**) the principle of source/sink completeness [☞ [31] and references therein]:

> ***Conclusion 11.8 (conjecture)*** *In consistent quantum models with gravity, (**1**) all symmetries are gauged, and (**2**) all sources/sinks (electric and magnetic) for all gauge fields are included and satisfy the appropriate generalization of the Dirac dual charge quantization condition [☞ Section 5.2.3 and relation (5.108)].*

This conclusion is so far the strictest known formulation of fundamental limitations on models in the string theoretical system, and so far has the status of a very strong *conjecture*: in spite of very strong indications, there is no rigorous proof (as yet)☝ [☞ also the discussion in the Polchinski–Smolin debate [429, 435]].

The early twenty-first century understanding of the so-called string theory – and especially with its M- and F-theory extensions – is by far not "just a theory." Just as classical and statistical mechanics are (axiomatic) theoretical systems that provide a conceptual and technical framework for addressing large classes of respective phenomena, so is "string theory" also a theoretical *system* and not a (single) theory. In fact, "string theory" includes *three* major theories: quantum theory, gauge theory and the theory of relativity, and is moreover the one known framework that unifies them in a coherent, cohesive and logically consistent fashion.

While all these considerations in no way oblige Nature to be describable within the string theoretical system, they do make it our best candidate, ever.

Discrete spacetime

The whole basis of the Democritean atomistic worldview relies on the idea that – everyday sensory experiences to the contrary – matter is not continuous, but consists of an immense number of very teeny *elementary* particles. The fundamental physics of the twentieth century, and foremost quantum physics, convinces us that all existent matter, including also the interaction fields, may be described in this same atomistic fashion:

Conclusion 11.9 *All "ingredients" of the Standard Model [☞ Section 7.3.3] and including the gravitational field [☞ Chapter 9] are represented by quantum fields, i.e., elementary* ***particles*** *[☞ the lexicon entry, in Appendix B.1, entries for* ***field (physical)*** *and* ***quantum****, as well as Footnote 17 on p. 196].*

Thus, there is no logical obstruction for spacetime to also be discrete. For example, from the worldsheet perspective of the string theoretical system, target spacetime is simply the space of values spanned (dynamically generated) by the coordinate fields such as $X^\mu(\xi)$ in the Polyakov action (11.10).

The idea that spacetime is (at least in some directions) not in fact a continuous, commutative *topological space* is not new [☞ e.g., [130, 92, 361] and references therein]. One of many and various possibilities is very close to the computational method that is used in so-called "lattice QCD" [☞ description on p. 230]. The results of recent experiments in the LHC installation at CERN seem to lend support to a variant of this idea [374, 14, 13]. Here, in the simplest *model*, spacetime literally has the lattice structure of a crystal, in the sense that it consists of discrete points through which all material objects pass as if those points are ordered at constant distances and in uniform directions, just like atoms in crystalline lattices. Note that this is a very radical idea where the *auxiliary* space (the one throughout which the points of the *true* spacetime are distributed in regularly periodic and uniform fashion as a crystalline lattice) is a purely fictitious structure. If we further assume that the points in this "crystalline" spacetime are ordered akin to Cartesian directions and that the distances between the points in different directions are significantly different,

$$L_z \gg L_y \gg L_x \gg L_{ct}, \tag{11.76}$$

then, denoting λ_{dB} the de Broglie wavelength, we have that:

1. to probes with $\lambda_{\text{dB}} > L_z$; spacetime appears continuous and 3+1 dimensional;
2. to probes with $L_z > \lambda_{\text{dB}} > L_y$ the whole spacetime appears to be a vertical stack (uniform sequence) of horizontal and continuous $(2+1)$-dimensional (surface) spacetimes;
3. to probes with $L_y > \lambda_{\text{dB}} > L_x$ the whole spacetime appears to be a two-directional stack of horizontal and continuous $(1+1)$-dimensional (linear) spacetimes;
4. to probes with $L_x > \lambda_{\text{dB}} > L_{ct}$ the whole spacetime appears to be a three-directional stack of $(0+1)$-dimensional space-like arrangement of points with (still) a continuous passage of time;
5. to probes with $L_{ct} > \lambda_{\text{dB}}$ the whole spacetime appears to be a four-directional stack of spacetime *points*, i.e., disconnected events.

Amusingly, such a discrete structure of spacetime is implied by the assumption that the space of conjugate momenta is compact. This concrete Cartesian "crystalline" structure is simply obtained, e.g., in the *momentum* representation of quantum mechanics where one imposes periodic conditions to the linear momenta,[36] which gives to the momentum space the geometry of a 3-dimensional torus, the radii proportional to the reciprocals of the distances L_x^{-1}, L_y^{-1} and L_z^{-1}. Evidently, more complicated (and for now ad hoc imposed) compact geometry of the 4-momentum space then implies a more complicated structure of the discrete spacetime.

11.4.3 Lessons of fundamental physics as a model of nature

The duality between different – and even different-dimensional – models reminds us that the picturesque formulation imagery at the foundation of a given model is merely a mental caricature and image – a crutch – just as the hydrogen atom is merely *represented/imagined* as a point-like

[36] Within non-relativistic quantum mechanics at least, it seems essentially contradictory to require periodicity of energy.

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

electron orbiting a point-like proton. The *real* atom is not two point-like charges orbiting each other, nor is the atom a charged standing wave undulating along a circle around an oppositely charged point-particle, nor is the atom a negatively charged cloud centered on a positively charged proton… Of course, such picturesque formulation imagery is very useful in describing the atom, in that it dictates the construction of a corresponding mathematical model, which is then used to "produce" predictions of the model – intending to compare those predictions with Nature.

The connection between the picturesque formulation imagery and the ultimate authority, Nature, is very indirect, and so then is the justification of the formulation imagery, however impressively picturesque it may be. It should thus come as no surprise that even very different formulation images may turn out to produce models that agree equally with Nature [☞ Definition 11.3 on p. 437, and the discussion after this definition; see also the recent work [222]].

The caution that "the map is not the territory" (Alfred Korzybsky) is perfectly in agreement with this lesson, and omits the cultural–historical and perhaps even religious connotations of the ancient Tao principle "The way you can go is not the real way. The name you can say is not the real name" [527].

11.4.4 Exercises for Section 11.4

✎ ***11.4.1*** *Verify the results (11.69).*

✎ ***11.4.2*** *Using the definitions (11.72), generalize the flux definitions (11.63a) and (11.64a) for a rank-r antisymmetric gauge potential. Show the conditions (11.73a) and (11.73b) to reproduce the same quantization condition for the product $q_e^{(r)} q_m^{(n-r-2)}$.*

11.5 Instead of an epilogue: unified theory of everything

The fluffiness of clouds and the babbling of a brook are examples of emergent phenomena, which is not within the domain of fundamental physics, but of the physics of collectives – and that is a relatively *new* and emerging discipline in physics. The subject matter here is precisely the regularity and circumstances wherein relatively simple basic rules and their theoretical systems may produce (by means of nonlinear and/or self-interactive coupling) very complex phenomena.

For example, the basic laws of chemical bonds are relatively simple and stem from elementary quantum mechanics [☞ Schrödinger's quotation on p. 13 and its discussion], but nevertheless produce a fantastically diverse palette of an uncounted number of chemical compounds, as well as different materials. These compounds and materials then, through their dynamics and interactions, produce complex structures and behaviors that can in no particular sense or case be simply and *completely* reduced to elementary quantum mechanics, although in a completely literal sense they stem – draw roots – from it.

This reminds us of the fact that the stability, functionality and beauty of a palace are not properties of its bricks, shingles and other materials of which the palace is built [☞ also the discussion in Section 1.1.4]. Similarly, neither is the evolutionary role, nor the mimicry function or the beauty of the complicated patterns on butterfly wings simply the "diffraction and coherent scattering of light," although this is the basic mechanism for the appearance of most colors in the often stunningly exquisite wings.

In a sense, akin to the term "epiphenomenon," this new discipline could be called *epiphysics*. On one hand, this new discipline would be concerned with phenomena that are *beyond* currently familiar physics, and on the other, the subject matter of this new discipline would still be the *Nature* (φύσις ≈ *physis*, Greek) of this next level of natural phenomena. However, it is important to keep in mind that the demarcation between fundamental physics and this *epiphysics* must be hazy; in the end, Nature is one [☞ Conclusion 11.2 on p. 409].

https://doi.org/10.1017/9781009291507.015 Published online by Cambridge University Press

The helix of learning has thus come full circle, and hopefully one floor higher: Bohr's thought, quoted in the Preface, on p. xi, resonates through the entire development of the fundamental physics of elementary particles, appears explicitly also in the discussion around Digression 1.1 on p. 9, then again in Section 8.3, and completely permeates Chapter 11 and especially Section 11.4. At any rate, during the twentieth century, the fundamental physics of elementary particles has been developing from a discipline in which one believed to have almost everything solved to a discipline that is bound to separate into at least two or three separate disciplines within physics [☞ Section 11.2], and possibly also into a discipline the subject matter of which is the *structure* of (theoretical) physics itself. Evidently, within this development, a theoretical system has been discovered within which there is hope of finding a description of the *fundamentals* of Nature, but this has, *en route*, instructed us very pointedly about the very nature of our understanding of Nature.