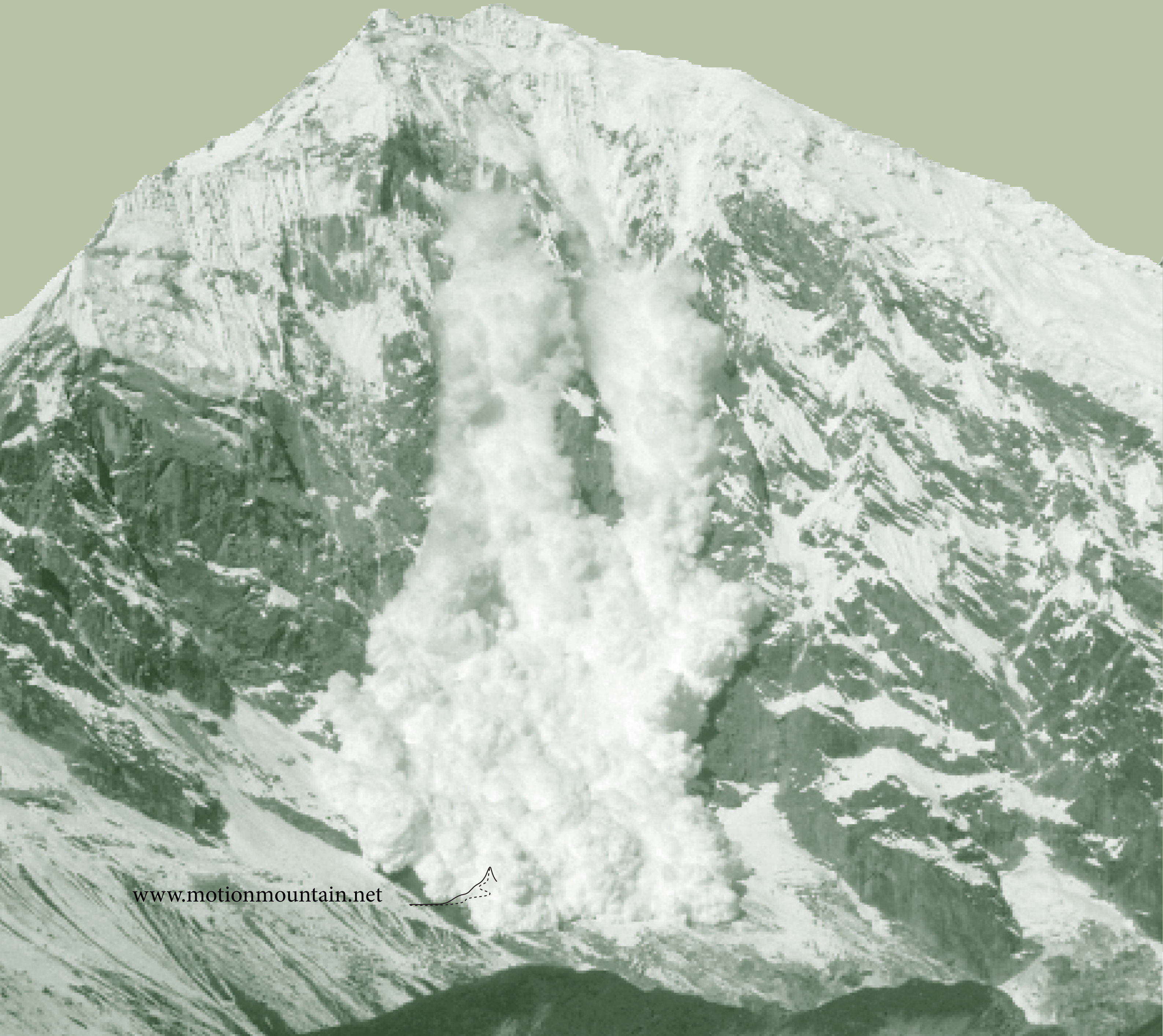
Christoph Schiller

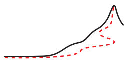
MOTION MOUNTAIN

THE ADVENTURE OF PHYSICS – VOL. VI

A SPECULATION ON UNIFICATION

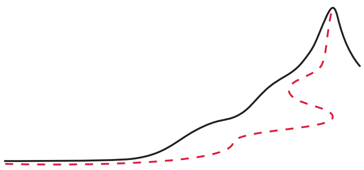


www.motionmountain.net



Christoph Schiller

Motion Mountain



The Adventure of Physics
Volume VI

A Speculation on Unification

Edition 24.26, available as free pdf at
www.motionmountain.net

Editio vicesima quarta.

Twenty-fourth edition.

To Britta, Esther and Justus Aaron

τῷ ἐμοὶ δαὶμονι

Die Menschen stärken, die Sachen klären.

PREFACE

This book is written for anybody who is intensely curious about nature and motion. Have you ever asked: Why do people, animals, things, images and empty space move? The answer leads to many adventures, and this book presents one of the best of them: the search for a precise, unified and final description of *all* motion.

The wish to describe *all* motion is a large endeavour. Fortunately, this large endeavour can be structured in the simple diagram shown in Figure 1. The *final* and *unified* description of motion, the topic of this book, corresponds to the highest point in the diagram. Searching for this final and unified description is an old quest. In the following, I briefly summarize its history and then present an intriguing, though *speculative* solution to the riddle. The approach is an unexpected result from a threefold aim that I have pursued since 1990, in the five previous volumes of this series: to present the basics of motion in a way that is up to date, captivating and simple. In retrospect, the aim for maximum simplicity has been central in deducing this speculation.

The search for the final, unified description of motion is a story of many surprises. For example, twentieth-century research has shown that there is a smallest distance in nature. Research has also shown that matter cannot be distinguished from empty space at those small distances. A last surprise dates from this century: particles and space are best described as made of *strands*, instead of little spheres or points. The present text explains how to reach these unexpected conclusions. In particular, quantum field theory, the standard model of particle physics, general relativity and cosmology are shown to follow from strands. The three gauge interactions, the three particle generations and the three dimensions of space turn out to be due to strands. In fact, all the open questions of twentieth-century physics about the foundations of motion, all the millennium issues, can be solved with the help of strands.

The ideas in this text, in full contrast to those of the five previous volumes, are speculative. While the previous volumes introduced, in an entertaining way, the *established* parts of physics, this volume presents, in the same entertaining and playful way, a *speculation* about unification. Nothing in this volume is established knowledge – yet.

The search for a final theory is one of the great adventures of life: it leads to the limits of thought. The search overthrows our thinking habits about nature. A change in thinking habits can produce fear, often hidden by anger. But by overcoming our fears we gain strength and serenity. Changing thinking habits thus requires courage, but it also produces intense and beautiful emotions. Enjoy them!

Munich, 3 February 2011.

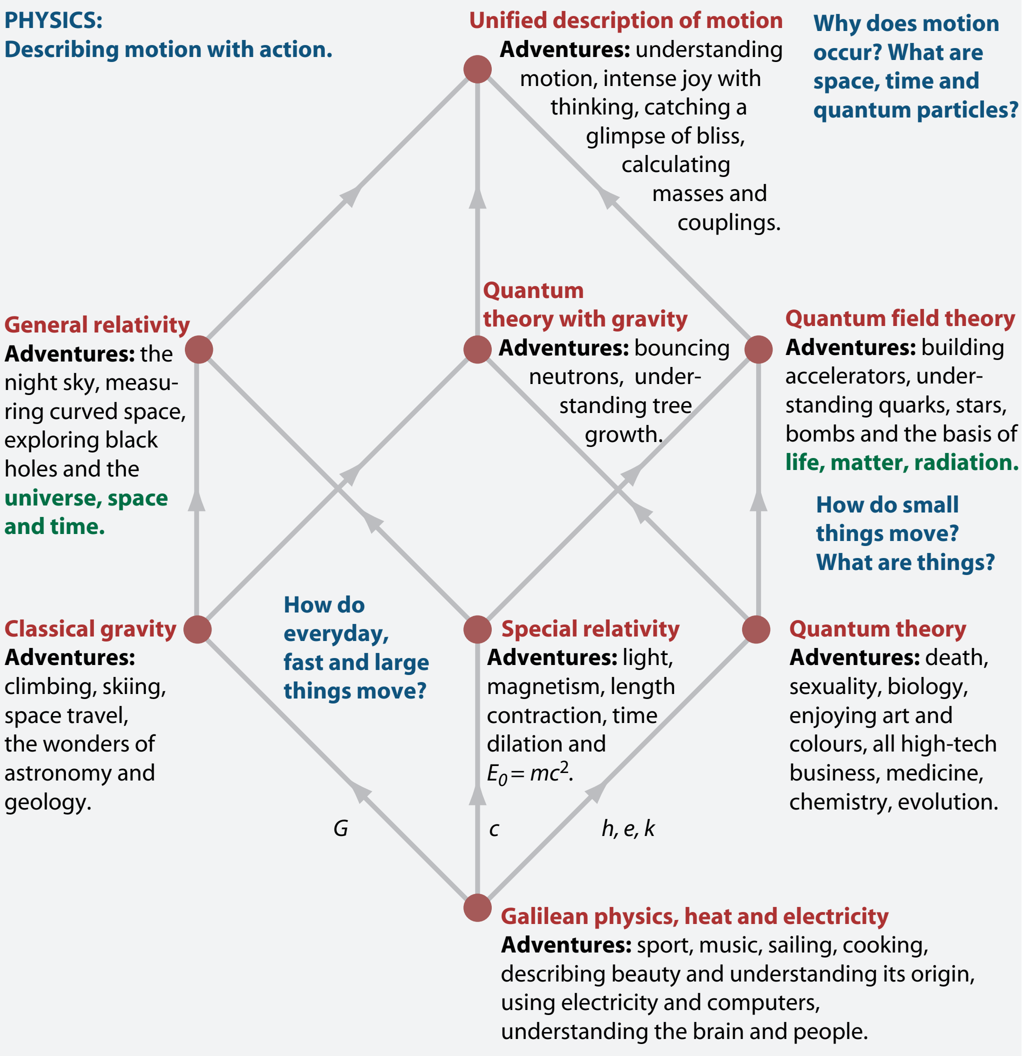


FIGURE 1 A complete map of physics: the connections are defined by the speed of light c, the gravitational constant G, the Planck constant h, the Boltzmann constant k and the elementary charge e.

USING THIS FILE

Text in green, as found in many marginal notes, marks a link that can be clicked in a pdf reader. Such green links are either bibliographic references, footnotes, cross references to other pages, challenge solutions, or pointers to websites.

Solutions and hints for *challenges* are given in the appendix. Challenges are classified as research level (r), difficult (d), standard student level (s) and easy (e). Challenges of type r, d or s for which no solution has yet been included in the text are marked (ny).

This sixth volume of the Motion Mountain series has been typeset in a way that printing the file in black and white gives the smallest possible reduction in reading pleasure.

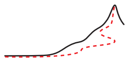
SUPPORT

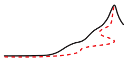
I would be delighted to receive an email from you at fb@motionmountain.net, especially on the following issues:

Challenge 1 s

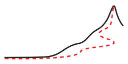
— What was missing or hard to follow and should be clarified?
— What should be corrected?

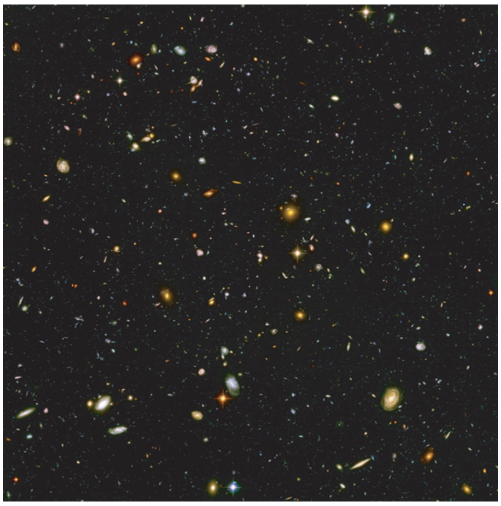
Alternatively, you can provide feedback online, on www.motionmountain.net/wiki. The feedback will be used to improve the next edition. On behalf of all readers, thank you in advance for your input. For a particularly useful contribution you will be mentioned – if you want – in the acknowledgements, receive a reward, or both. But above all, enjoy the reading!





CONTENTS





A SPECULATION ON UNIFICATION

Where, through the combination of
quantum mechanics and general relativity,
the top of Motion Mountain is reached,
and it is discovered
that vacuum is indistinguishable from matter,
that there is little difference between the large and the small,
that nature can be described by strands,
that particles can be modelled as tangles,
that interactions appear naturally,
and that a complete description of motion is possible.

CHAPTER 1

FROM MILLENNIUM PHYSICS TO UNIFICATION

Look at what happens around us. A child who smiles, a nightingale that sings, a lily that opens: all move. Even an immobile shadow is due to moving light. Every star owes its formation and its shine to various kinds of motion. The darkness of the night sky is due to motion: it results from the expansion of empty space.* Finally, human creativity is due to the motion of molecules, ions and electrons in the brain. Is there a common language for all these observations?

Is there a *unified* and *precise* way to describe all motion? Is everything that moves, from people to planets, from light to empty space, made of the same constituents? What is the origin of motion? Answering these questions is the topic of the present text.

Answering questions about motion with precision defines the subject of *physics*. Over the centuries, a huge number of precise observations about motion have been collected. We now know how electric signals move in the brain, how insects fly, how colours appear, how the stars formed, how life evolved, and much more. We use our knowledge about motion to look into the human body and heal illnesses; we use our knowledge about motion to build electronics, communicate over large distances, and work for peace; we use our knowledge about motion to secure life against many of nature's dangers, including droughts and storms. Physics, the science of motion, has shown time after time that knowledge about motion is fascinating and useful.

At the end of the last millennium, humans were able to describe *all* observed motion with high precision. This description can be summarized in a few statements.

— In nature, motion takes place in three dimensions of space and is described by the least action principle. Action is a physical quantity that describes how much change occurs in a process. The least action principle states: *motion minimizes change*. Among others, the least change principle implies that motion is predictable, that energy is conserved and that growth and evolution are natural processes, as is observed. Ref. 1, Ref. 3

— In nature, there is an invariant maximum energy speed, the speed of light c. This invariant maximum implies *special relativity*. Among others, it implies that mass and energy are equivalent, as is observed. Ref. 2

— In nature, there is an invariant highest momentum flow, the Planck force $c^4/4G$. This invariant maximum implies *general relativity*. Among others, it implies that things fall and that empty space curves and moves, as is observed. Ref. 2

---

* The photograph on page 16 shows an extremely distant, thus extremely young, part of the universe, with its large number of galaxies in front of the black night sky (courtesy NASA).

— The evolution of the universe is described by the cosmological constant Λ. It determines the largest distance and the largest age that can presently be observed. Ref. 2

— In nature, there is a non-zero, invariant smallest change, the quantum of action $\hbar$. This invariant value implies *quantum theory*. Among others, it explains what life and death are, why they exist and how we enjoy the world. Ref. 4

— In nature, matter and radiation consist of quantum particles. Matter consists of *fermions*: six quarks, three charged leptons, three neutrinos and their antiparticles. Radiation consists of *bosons*: the photon, three intermediate weak vector bosons and eight gluons. Fermions and bosons move and can transform into each other. The transformations are described by the electromagnetic interaction, the weak nuclear interaction and the strong nuclear interaction. Together with the masses, quantum numbers, mixing angles and couplings, these transformation rules form the so-called *standard model of particle physics*. Among others, the standard model explains how lightning forms, how colours appear, and how the atoms in our bodies came to be. Ref. 4

These statements, the *millennium description of physics*, describe everything known in the year 2000 about motion. These statements describe the motion of people, animals, plants, objects, light, radiation, stars, empty space and the universe. Only a surprisingly small set of observations does not yet follow from these statements. A famous example is the nature of dark matter. We do not know yet what dark matter is. Another example is the way thinking forms in our brain. We do not know yet in detail how thinking follows from the above statements, though we do know that thinking is not in contrast with them. In the case of dark matter this is not so clear: dark matter could be in contrast with the millennium description of motion.

In other words, even though the millennium description of physics is precise and successful, there are some open issues. In particular, the last statement given above, on the standard model, is not as simple as the preceding ones. How is the standard model related to the preceding statements? Why are there *three* interactions, *twelve* elementary fermions, *twelve* elementary bosons and *three* dimensions? And why is there motion anyway? These issues form the quest for unification, phrased in concrete terms.

The complete list of all those *fundamental* issues about motion that were *unexplained* in the year 2000 make up only a short table. We call them the *millennium issues*. The quest for unification – and the topic of this text – is their solution. A *final theory of motion* is a theory that solves these issues.

TABLE 1 The millennium list: *everything* particle physics and general relativity *cannot* explain; thus, also the list of the *only* experimental data available to test the final, unified description of motion.

OBSERVABLE	PROPERTY UNEXPLAINED IN THE YEAR 2000
Local quantities, from quantum field theory: particle properties	
$\alpha = 1/137.036(1)$	the low energy value of the electromagnetic coupling constant
α_w or θ_w	the low energy value of the weak coupling constant or the value of the weak mixing angle
α_s	the value of the strong coupling constant at one specific energy value
m_q	the values of the 6 quark masses

TABLE 1 (Continued) *Everything* the standard model and general relativity *cannot* explain.

OBSERVABLE	PROPERTY UNEXPLAINED IN THE YEAR 2000
m_l	the values of 6 lepton masses
m_W	the value of the mass of the W vector boson
m_H	the value of the mass of the scalar Higgs boson
$\theta_{12}, \theta_{13}, \theta_{23}$	the value of the three quark mixing angles
δ	the value of the CP violating phase for quarks
$\theta^{\nu}_{12}, \theta^{\nu}_{13}, \theta^{\nu}_{23}$	the value of the three neutrino mixing angles
$\delta^{\nu}, \alpha_1, \alpha_2$	the value of the three CP violating phases for neutrinos
$3 \cdot 4$	the number of fermion generations and of particles in each generation
J, P, C, etc.	the origin of all quantum numbers of each fermion and each boson
Local mathematical structures, from quantum field theory	
$c, \hbar, k$	the origin of the invariant Planck units of quantum field theory
$3 + 1$	the number of dimensions of physical space and time
SO(3,1)	the origin of Poincaré symmetry, i.e., of spin, position, energy, momentum
$S(n)$	the origin of particle identity, i.e., of permutation symmetry
Gauge symmetry	the origin of the gauge groups, in particular:
U(1)	the origin of the electromagnetic gauge group, i.e., of the quantization of electric charge, as well as the vanishing of magnetic charge
SU(2)	the origin of weak interaction gauge group, its breaking and P violation
SU(3)	the origin of strong interaction gauge group and its CP conservation
Ren. group	the origin of renormalization properties
$\delta W = 0$	the origin of wave functions and the least action principle in quantum theory
$W = \int L_{SM}\, dt$	the origin of the Lagrangian of the standard model of particle physics
Global quantities, from general relativity: vacuum and energy properties	
0	the observed flatness, i.e., vanishing curvature, of the universe
$1.2(1) \cdot 10^{26}$ m	the distance of the horizon, i.e., the 'size' of the universe (if it makes sense)
$\rho_{de} = \Lambda c^4/(8\pi G)$ $\approx 0.5\,\mathrm{nJ/m^3}$	the value and nature of the observed vacuum energy density, dark energy or cosmological constant
$(5 \pm 4) \cdot 10^{79}$	the number of baryons in the universe (if it makes sense), i.e., the average visible matter density in the universe
$f_0(1, ..., c.\ 10^{90})$	the initial conditions for *c.* 10^{90} particle fields in the universe (if or as long as they make sense), including the homogeneity and isotropy of matter distribution, and the density fluctuations at the origin of galaxies
ρ_{dm}	the density and nature of dark matter
Global mathematical structures, from general relativity	
c, G	the origin of the invariant Planck units of general relativity
$\delta \int L_{GR} dt = 0$	the origin of the least action principle and the Lagrangian of general relativity
$R \times S^3$	the observed topology of the universe

Against a final theory

A fixed list of arguments are repeated regularly against the search for a final, unified theory of motion. Reaching the final theory and enjoying the adventure is only possible if these arguments are known – and then put gently aside.

— It is regularly said that a final theory cannot exist because nature is infinite and mysteries will always remain. But this statement is wrong. First, nature is not infinite. Second, even if it were infinite, knowing and describing everything would still be possible. Third, even if knowing and describing everything would be impossible, and if mysteries would remain, a final theory remains possible. A final theory is not useful for every issue of everyday life, such as choosing your dish on a menu or your future profession. A final theory is simply a full description of the foundations of motion: the final theory combines and explains particle physics and general relativity.

— It is sometimes argued that a final theory cannot exist due to Gödel's incompleteness theorem or due to computational irreducibility. However, in such arguments, both theorems are applied to domains were they are not valid. The reasoning is thus wrong.

— Some state that it is not clear whether a final theory exists at all. This is wrong. Physical theories are ways to *talk* about nature, and for the final theory we only have to search for those concepts that enable us to talk with precision about *all* of motion. Because we are looking for a way to talk, we know that the final theory must exist. And searching for it is fascinating and exciting, as everybody busy with this adventure will confirm.

Ref. 5

— Some claim that the search for a final theory is a reductionist endeavour and cannot lead to success, because reductionism is flawed. This claim is wrong on three counts. First, it is not clear whether the search is a reductionist endeavour, as will become clear later on. Second, there is no evidence that reductionism is flawed. Third, even if it were, no reason not to pursue the quest would follow. The claim in fact invites to search with a larger scope than was done in the past decades – an advice that will turn out to be spot on.

Ref. 6
Vol. IV, page 120

— Some argue that searching for a final theory makes no sense as long as the measurement problem of quantum theory is not solved, or consciousness is not understood, or the origin of life is not understood. Now, the measurement problem is solved by decoherence, and in order to combine particle physics with general relativity, understanding the details of consciousness or of the origin of life is not required. Neither is understanding or solving marriage problems required – though this might help.

— Some people claim that searching for a final theory is a sign of foolishness or a sin of pride. Such small and envious minds should simply be ignored; the nastier specimens might deserve to be ridiculed. The quest is the search for the solution to a riddle.

Ref. 7

— Some believe that understanding the final theory means to read the mind of god, or to think like god, or to be like god. This is false, as any expert on god will confirm. In fact, solving a riddle or reading a physics textbook does not transform people into gods. This is unfortunate, as such an effect would provide excellent advertising.

— Some fear that knowing the final theory yields immense power that harbours huge dangers of misuse, in short, that knowing the final theory might change people into

Ref. 8 devils. However, this fear is purely imaginary; it only describes the fantasies of the person that is talking. Indeed, the millennium description of physics is already quite near to the final theory, and nothing to be afraid of has happened. Sadly, another great advertising opportunity is eliminated.

- Some people object that various researchers in the past have thought to have found the final theory, but were mistaken, and that many great minds tried to find a final theory, but had no success. That is true. Some failed because they lacked the necessary tools for a successful search, others because they lost contact with reality, and still others because they were led astray by prejudices that limited their progress. We just have to avoid these mistakes.

In short, we can reach the final unified theory – which we symbolically place at the top of Motion Mountain – only if we are not burdened with ideological or emotional baggage. The goal we have set requires *extreme thinking*, i.e., thinking up to the limits. After all, unification is the precise description of *all* motion. Therefore, unification is a riddle; and searching for it is a pastime. A riddle is best approached with the lightness that is intrinsic to playing. Life is short: we should play whenever we can. Let's start. Ref. 9

WHAT WENT WRONG IN THE PAST

The twentieth century was the golden age of physics. Searching for the final theory, researchers explored candidates such as grand unified theories, supersymmetry, and various other options to be described later on. Vol. V, page 196 But all these candidates were either falsified by experiment or, worse, unrelated to experiment. In other words, despite a large number of physicists working on the problem, despite the availability of extensive experimental data, and despite several decades of research, no final theory was found. Why?

During the twentieth century, many successful descriptions of nature were deformed into dogmatic beliefs about unification. Here are the main examples, with some of their best known proponents:

- 'Unification requires generalization of existing theories' (almost everybody).
- 'Unification is independent of Planck's natural units' (almost everybody).
- 'Unification requires axiomatization' (David Hilbert).
- 'Unification requires evolution equations' (Albert Einstein, Werner Heisenberg).
- 'Unification requires space to be a manifold' (Albert Einstein).
- 'Unification requires searching for beauty' (Paul Dirac).
- 'Unification requires more dimensions of space' (Theodor Kaluza).
- 'Unification requires finding higher symmetries' (Abdus Salam).
- 'Unification requires Higgs bosons' (Steven Weinberg).
- 'Unification requires additional elementary particles' (Steven Weinberg).
- 'Unification requires supersymmetry' (Steven Weinberg).
- 'Unification requires complicated mathematics' (Edward Witten).
- 'Unification requires solving huge conceptual difficulties' (Edward Witten).
- 'Unification is only for a selected few' (many).

All these dogmas appeared in the same way: famous scholars – in fact many more than those given above – explained the idea that guided their past successes, and then they and most other researchers started to believe more the guiding idea than the result itself.

In fact, all the beliefs just mentioned can be seen as special cases of the first one. And like the first one, as we will discover in the following, they are all wrong.

How to find the final theory of motion

We have a riddle to solve: we want to describe precisely all motion and discover its origin. In order to do this, we need to find a final theory that solves and explains each open issue given in the *millennium list*. To find the final theory, we first simplify quantum theory and gravitation as much as possible, explore what happens when the two are combined, and deduce the *requirement list* that any final theory must fulfil. Then we deduce the simplest possible model that fulfils the requirements; we check the properties of the model against every experiment performed so far and against every open issue from the millennium list. Discovering that there are no disagreements, no points left open and no possible alternatives, we know that we have found the final theory. We thus end our adventure with a *list of testable predictions*.

In short, three lists structure our quest for a final theory: the millennium list of open issues, the list of requirements, and the list of testable predictions. To get from one list to the next, we proceed along the following legs.

1. We first simplify modern physics. Twentieth century physics deduced several *invariant* properties of motion. These invariants, such as the speed of light or the quantum of action, are called *Planck units*. The invariant Planck units allow motion to be measured. Above all, these invariants are also found to be *limit values*, valid for every example of motion.
2. Combining quantum theory and general relativity, we discover that at the Planck limits, the universe, space and particles are *not described by points*. We find that as long as we use points to describe particles and space, and as long as we use sets and elements to describe nature, a unified description of motion is impossible.
3. The combination of quantum theory and general relativity teaches us that space and particles have *common constituents*.
4. By exploring black holes, spin, and the limits of quantum theory and gravity, we discover that the common constituents of space and particles are fluctuating, extended, without ends, and one-dimensional: the common constituents of space and particles are *fluctuating strands*.
5. We discover that we cannot think or talk without continuity. We need a *background* to describe nature. We conclude that to talk about motion, we have to combine continuity and non-continuity in an appropriate way. This is achieved by imagining that fluctuating strands move in a continuous three-dimensional *background*.

At this point, after the first half of our adventure, we have obtained an extensive *requirement list* for the final theory. This list allows us to proceed rapidly to our goal, without being led astray. Page 149

6. We discover a simple fundamental principle that explains how the maximum speed c, the minimum action $\hbar$, the maximum force $c^4/4G$ and the cosmological constant Λ follow from strands. We also discover how to deduce quantum theory, relativity and cosmology from strands.

7. We discover that strands naturally yield the existence of three spatial dimensions, flat and curved space, black holes, the cosmological horizon, fermions and bosons. We find that all known physical systems are made from strands. Also the process of measurement and all properties of the background result from strands.
8. We discover that fermions emit and absorb bosons and that they do so with exactly those properties that are observed for the electromagnetic, the weak and the strong nuclear interaction. In short, the *three known gauge interactions* – and their parity conservation or violation – follow from strands. In addition, we discover that other interactions do not exist.
9. We discover that strands naturally yield the known elementary fermions and bosons, grouped in *three generations*, with all the properties that are observed. Other elementary particles do not exist. We thus recover the standard model of elementary particles.
10. We discover that the fundamental principle solves all the issues listed in the table of unexplained properties, and that all properties deduced from strands agree with experiment. Therefore, an extensive *list of testable predictions* can be given. They will all be tested – by experiment or by calculation – in the coming years.

Page 327

11. We discover that motion is the observation of crossing switches due to strand fluctuations. Motion is an inescapable consequence of observation: motion is an experience that we make because we are a small, approximate part of a large whole.

At the end of this path, we will thus have unravelled the mystery of motion. It is a truly special adventure. **But be warned: almost all of the story presented here is still speculative, and thus open to question.** With almost every sentence you will find at least one physicist who disagrees. That makes the adventure even more fun.

Chapter 2

PHYSICS IN LIMIT STATEMENTS

Twentieth century physics deduced several *invariant* properties of motion. These invariants, such as the speed of light or the quantum of action, define the so-called Planck units. The invariant Planck units are important for two reasons: first, they allow motion to be measured; second, the invariants are *limit values*. In fact, the Planck units provide bounds for all observables.

The main lesson of modern physics is thus the following: When we simplify physics as much as possible, we discover that *nature limits the possibilities of motion*. Such limits lie at the origin of special relativity, of general relativity and of quantum theory. In fact, we will see that nature limits *every* aspect of motion. Exploring the limits of motion will allow us to deduce several astonishing conclusions. These conclusions contradict all that we learned about nature so far.

SIMPLIFYING PHYSICS AS MUCH AS POSSIBLE

At dinner parties, physicists are regularly asked to summarize physics in a few sentences. It is useful to have a few simple statements ready to answer such a request. Such statements are not only useful to make other people think; they are also useful in our quest for the final theory. Here they are.

Classical physics in one statement

Everyday motion is described by Galilean physics. It consists of only one statement: all *motion minimizes change*. In nature, change is measured by physical action W. More precisely, change is measured by the average difference between kinetic and potential energy. In other words, motion obeys the so-called *least action principle*, written as

$$\delta W = 0 \text{ , where } \quad W = \int (T - U)\, \mathrm{d}t \ . \tag{1}$$

This statement determines the effort we need to move or throw stones, and explains why cars need petrol or people need food. In other terms, *nature is as lazy as possible*. The natural laziness of everyday motion remains valid throughout modern physics, provided a few limit statements are added.

SPECIAL RELATIVITY IN ONE STATEMENT

The step from everyday, or Galilean, physics to special relativity can be summarized in a single limit statement on motion. It was popularized by Hendrik Antoon Lorentz: *There is a maximum energy speed in nature.* For all physical systems and all observers, the local energy speed v is limited by the speed of light c: Ref. 10

$$v \leqslant c \ . \tag{2}$$

All results peculiar to special relativity follow from this principle. A few well-known facts set the framework for the discussion that follows. The speed v is less than or equal to the speed of light c for *all* physical systems;* in particular, this limit is valid both for composite systems and for elementary particles.

The energy speed limit is an *invariant*: the energy speed limit is valid for *all* observers. In this context it is essential to note that any observer must be a physical system, and must be *close* to the moving energy. Vol. II, page 89

The speed limit c is realized by *massless* particles and systems; in particular, it is realized by electromagnetic waves. For matter systems, the speed is always below c.

Only a maximum energy speed ensures that cause and effect can be distinguished in nature, or that sequences of observations can be defined. The opposite hypothesis, that energy speeds greater than c are possible, which implies the existence of (real) tachyons, has been explored and tested in great detail; it leads to numerous conflicts with observations. Tachyons do not exist.

The maximum energy speed forces us to use the concept of *space-time* to describe nature, because the existence of a maximum energy speed implies that space and time *mix*. It also implies observer-dependent time and space coordinates, length contraction, time dilation, mass–energy equivalence, horizons for accelerated observers, and all the other effects that characterize special relativity. Only a maximum speed leads to the principle of maximum ageing that governs special relativity; and only this principle leads to the principle of least action at low speeds. In addition, only with a finite speed limit is it possible to define a *unit* of speed that is valid at all places and at all times. If there were no global speed limit, there could be no natural measurement standard for speed, independent of all interactions; speed would not then be a measurable quantity. Vol. II, page 25

Special relativity also limits the size of systems – whether composite or elementary. Indeed, the limit speed implies that acceleration a and size l cannot be increased independently without bounds, because the two ends of a system must not interpenetrate. The most important case concerns massive systems, for which we have

$$l \leqslant \frac{c^2}{a} \ . \tag{3}$$

This size limit is induced by the speed of light c; it is also valid for the *displacement* d of

---

* A *physical system* is a region of space-time containing mass–energy, the location of which can be followed over time and which interacts incoherently with its environment. The speed of a physical system is thus an *energy speed*. The definition of physical system excludes images, geometrical points or incomplete, entangled situations.

a system, if the acceleration measured by an external observer is used. Finally, the speed limit implies a relativistic 'indeterminacy relation'

$$\Delta l \, \Delta a \leqslant c^2 \tag{4}$$

for the length and acceleration indeterminacies. You may wish to take a minute to deduce this relation from the time–frequency indeterminacy. All this is standard knowledge. Challenge 2 s

QUANTUM THEORY IN ONE STATEMENT

The difference between Galilean physics and quantum theory can be summarized in a single statement on motion, due to Niels Bohr: *There is a minimum action in nature.* For all physical systems and all observers, the action W obeys Ref. 11 Vol. IV, page 14

$$W \geqslant \hbar \, . \tag{5}$$

The Planck constant $\hbar$ is the smallest observable action or change of angular momentum. This statement is valid for all systems, thus both for composite and elementary systems. The principle contains all of quantum theory. We call it the *principle of non-zero action*, in order avoid confusion with the principle of least action.

The non-zero action limit $\hbar$ is an *invariant*: it is valid with the same numerical value for *all* observers. Again, any observer must be a physical system.

The action limit is realized by many physical processes, from the absorption of light to the flip of a spin 1/2 particle. More precisely, the action limit is realized by *microscopic* systems that are made of a single particle.

The non-zero action limit is stated less frequently than the speed limit. It starts from the usual definition of the action, $W = \int (T - U) \, \mathrm{d}t$, and states that between two observations performed at times t and $t + \Delta t$, even if the evolution of a system is not known, the measured action is at least $\hbar$. Physical action measures the change in the state of a physical system. Thus there is always a minimum change of state between two different observations of a system.* The non-zero action limit expresses the fundamental fuzziness of nature at a microscopic scale.

It can easily be checked that no observation – whether of photons, electrons or macroscopic systems – gives a smaller action than the value $\hbar$. The non-zero action limit has been verified for fermions, bosons, laser beams and matter systems, and for any combination of these. The opposite hypothesis, implying the existence of arbitrary small change, has been explored in detail: Einstein's long discussion with Bohr, for example, can be seen as a repeated attempt by Einstein to find experiments that would make it possible to measure arbitrarily small changes or action values in nature. In every case, Bohr found that this could not be achieved. All subsequent attempts were equally unsuccessful.

Ref. 12 The principle of non-zero action can be used to deduce the indeterminacy relation, the tunnelling effect, entanglement, permutation symmetry, the appearance of probabilities in quantum theory, the information-theoretic formulation of quantum theory, and the existence of elementary particle reactions. It implies that in quantum theory, the three

---

* For systems that seem constant in time, such as a spinning particle or a system showing the quantum Zeno effect, finding this minimum change is tricky. Enjoy the challenge.

concepts of state, measurement operation, and measurement result need to be distinguished from each other; this is done by means of a so-called *Hilbert space*. The non-zero action limit is also the foundation of Einstein–Brillouin–Keller quantization. Ref. 13

The existence of a non-zero action limit has been known from the very beginning of quantum theory. It is at the basis of – and completely equivalent to – all the standard formulations of quantum theory, including the many-path and the information-theoretic formulations.

We also note that only a non-zero action limit makes it possible to define a *unit* of action. If there were no action limit, there could be no natural measurement standard for action: action would not then be a measurable quantity.

The upper action and speed bounds $W \leqslant pd \leqslant mcd$ for any physical system, together with the quantum of action, imply a limit on the displacement d of a system between any two observations:

$$d \geqslant \frac{\hbar}{mc} . \qquad (6)$$

In other words, the (reduced) Compton wavelength of quantum theory appears as the lower limit on the displacement of a system, whenever gravity plays no role. Since the quantum displacement limit applies in particular to an elementary system, it also applies to the *size* of a *composite* system. However, the limit is *not* valid for the sizes of elementary particles. Challenge 3 e

The limit on action also implies Heisenberg's well-known indeterminacy relation for the displacement d and momentum p of physical systems: Vol. IV, page 22

$$\Delta d \; \Delta p \geqslant \frac{\hbar}{2} . \qquad (7)$$

This relation is valid for both massless and massive systems. All this is textbook knowledge.

Thermodynamics in one statement

Thermodynamics can also be summarized in a single statement about motion: *There is a smallest entropy in nature.*

$$S \geqslant k . \qquad (8)$$

The entropy S is limited by the Boltzmann constant k. This result is almost 100 years old; it was stated most clearly by Leo Szilard. All of thermodynamics can be deduced from this relation, together with the quantum of action. Ref. 14

The entropy limit is an *invariant*: it is valid for *all* observers. Again, any observer must be a physical system.

The entropy limit is realized only by physical systems made of a single particle. In other words, the entropy limit is again realized only by *microscopic* systems. Therefore the entropy limit provides the same length limit for physical systems as the action limit.

Like the other limit statements we have examined, the entropy limit can also be phrased as a indeterminacy relation between temperature T and energy U:

$$\Delta\frac{1}{T}\,\Delta U \geqslant \frac{k}{2}\,. \tag{9}$$

Ref. 15 This relation was first given by Bohr and discussed by Heisenberg and many others.

GENERAL RELATIVITY IN ONE STATEMENT

Less well known is the possibility of summarizing the step from universal gravity to general relativity in a single statement on motion: *There is a maximum force or power in nature.*

For all physical systems and all observers, force F and power P are limited by

$$F \leqslant \frac{c^4}{4G} = 3.0 \cdot 10^{43}\,\mathrm{N} \quad \text{and} \quad P \leqslant \frac{c^5}{4G} = 9.1 \cdot 10^{51}\,\mathrm{W}\,. \tag{10}$$

These limit statements contain both the speed of light c and the gravitational constant G; they thus qualify as statements about relativistic gravitation.

Force is change of momentum; power is change of energy. Since momentum and energy are conserved, force and power are the flow of momentum and energy *through a surface*. Force and power, like electric current, describe the change in time of conserved quantity. For electric current, the conserved quantity is charge, for force, it is momentum, for power, it is energy. In other words, like current, also force is a flow across a surface. Vol. I, page 185 This is a simple consequence of the continuity equation. As a consequence, every discussion of maximum force implies a clarification of the underlying surface.

Both the force and the power limits state that the flow of momentum or of energy through any *physical surface* (a term defined below) of any size, for any observer, in any coordinate system, never exceeds the limit value. In particular, the force and power limits are realized only at *horizons*. In all other situations, the observed values are strictly smaller than the maximum values.

The force and power limit values are *invariants*: they are valid for *all* observers and for all interactions. Again, any observer must be a physical system and it must be located on or near the surface used to define the flow of momentum or energy. Vol. II, page 97

The value of the force limit is the energy of a Schwarzschild black hole divided by its diameter; here the 'diameter' is defined as the circumference divided by π. The power limit is realized when such a black hole is radiated away in the time that light takes to travel along a length corresponding to the diameter.

The value of the maximum force, as well as being the mass–energy of a black hole divided by its diameter, is also the surface gravity of a black hole times its mass. Thus the force limit means that no physical system of a given mass can be concentrated in a region of space-time smaller than a (non-rotating) black hole of that mass. In fact, the Challenge 4 e mass–energy concentration limit can easily be transformed algebraically into the force limit: they are equivalent.

It is easily checked that the maximum force limit is valid for all systems observed in nature, whether they are microscopic, macroscopic or astrophysical. Neither the 'gravitational force' (as long as it is operationally defined) nor the electromagnetic or nuclear interactions are ever found to exceed this limit. Challenge 5 e

But is it possible to *imagine* a system that exceeds the limit? An extensive discussion shows that this is impossible, if the size of observers or test masses is taken into account. Vol. II, page 95 Boosts do not help, as the transformed force value never exceeds the proper or comoving force value. Also changing to an accelerated observer does not help, because for high accelerations a, horizons appear at distance a/c^2, and a mass m has a minimum diameter given by $l \geqslant 4Gm/c^2$.

Because of the lack of nearby black holes or horizons, neither the force limit nor the power limit are realized in any physical system found so far, neither at everyday length scales, nor in the microscopic world or in astrophysical systems. But even in Gedanken experiments the limits cannot be exceeded, as long as the sizes of observers or of test masses are taken into account. Vol. II, page 95 All apparent exceptions assume point particles or observers, which are, however, unphysical and do not exist in general relativity.

The formulation of general relativity as a consequence of a maximum force is not common; in fact, it seems that it was only discovered 80 years after the theory of general relativity had first been proposed. Ref. 16

Deducing general relativity*

In order to elevate the force or power limit to a principle of nature, we have to show that, just as special relativity follows from the maximum speed, so general relativity follows from the maximum force.

Ref. 17 The maximum force and the maximum power are only realized at horizons. Horizons are regions of space-time where the curvature is so high that it limits the possibility of observation. The name 'horizon' is due to an analogy with the usual horizon of everyday life, which also limits the distance to which one can see. However, in general relativity horizons are *surfaces*, not lines. In fact, we can *define* the concept of horizon in general relativity as a region of maximum force; it is then easy to prove that a horizon is always a two-dimensional surface, and that it is essentially black (except for quantum effects).

The connection between horizons and the maximum force or power allows us to deduce the field equations in a simple way. First, there is always a flow of energy at a horizon. Horizons cannot be planes, since an infinitely extended plane would imply an infinite energy flow. To characterize the finite extension of a given horizon, we use its radius R and its total area A.

The energy flow across a horizon is characterized by an energy E and a proper length L of the energy pulse. When such an energy pulse flows perpendicularly across a horizon, the momentum change $\mathrm{d}p/\mathrm{d}t = F$ is given by

$$F = \frac{E}{L} \,. \qquad (11)$$

Since we are at a horizon, we need to insert the maximum possible values. In terms of

---

* This section can be skipped at first reading.

the horizon area A and radius R, we can rewrite the limit case as

$$\frac{c^4}{4G} = \frac{E}{A} 4\pi R^2 \frac{1}{L} \tag{12}$$

where we have introduced the maximum force and the maximum possible area $4\pi R^2$ of a horizon of (maximum local) radius R. The ratio E/A is the energy per unit area flowing across the horizon.

Horizons are often characterized by the so-called surface gravity a instead of the radius R. In the limit case, two are related by $a = c^2/2R$. This leads to

$$E = \frac{1}{4\pi G}\, a^2\, A\, L\ . \tag{13}$$

Ref. 18 Special relativity shows that at horizons the product aL of proper length and acceleration is limited by the value $c^2/2$. This leads to the central relation for the energy flow at horizons:

$$E = \frac{c^2}{8\pi G}\, a\, A\ . \tag{14}$$

This *horizon relation* makes three points. First, the energy flowing across a horizon is limited. Secondly, this energy is proportional to the area of the horizon. Thirdly, the energy flow is proportional to the surface gravity. These three points are fundamental, and characteristic, statements of general relativity. (We also note that due to the limit property of horizons, the energy flow *towards* the horizon just outside it, the energy flow *across* a horizon, and the energy *inside* a horizon are all the same.)

Taking differentials, the horizon relation can be rewritten as

$$\delta E = \frac{c^2}{8\pi G}\, a\, \delta A\ . \tag{15}$$

In this form, the relation between energy and area can be applied to general horizons, in particular those that are irregularly curved or time-dependent.*

Ref. 19 In a well-known paper, Jacobson has given a beautiful proof of a simple connection: if energy flow is proportional to horizon area for all observers and all horizons, and if the proportionality constant is the correct one, then general relativity follows. To see the connection to general relativity, we generalize the horizon relation (15) to general coordinate systems and general directions of energy flow.

---

* The horizon relation (15) is well known, though with different names for the observables. Since no communication is possible across a horizon, the detailed fate of energy flowing across a horizon is also unknown. Energy whose detailed fate is unknown is often called *heat*, and abbreviated Q. The horizon relation (15) therefore states that the heat flowing through a horizon is proportional to the horizon area. When quantum theory is introduced into the discussion, the area of a horizon can be called 'entropy' S and its surface gravity can be called 'temperature' T; relation (15) can then be rewritten as $\delta Q = T \delta S$. However, this translation of relation (15), which requires the quantum of action, is unnecessary here. We only cite it to show the relation between horizon behaviour and quantum gravity.

The proof uses tensor notation. We introduce the general surface element $\mathrm{d}\Sigma$ and the local boost Killing vector field k that generates the horizon (with suitable norm). We then rewrite the left-hand side of relation (15) as

$$\delta E = \int T_{ab} k^a \,\mathrm{d}\Sigma^b \,, \tag{16}$$

where T_{ab} is the energy–momentum tensor. This is valid in arbitrary coordinate systems and for arbitrary energy flow directions. Jacobson's main result is that the right-hand side of the horizon relation (15) can be rewritten, using the (purely geometric) Raychaudhuri equation, as

$$a \,\delta A = c^2 \int R_{ab} k^a \,\mathrm{d}\Sigma^b \,, \tag{17}$$

where R_{ab} is the Ricci tensor describing space-time curvature.

Combining these two steps, we find that the energy–area horizon relation (15) can be rewritten as

$$\int T_{ab} k^a \,\mathrm{d}\Sigma^b = \frac{c^4}{8\pi G} \int R_{ab} k^a \,\mathrm{d}\Sigma^b \,. \tag{18}$$

Jacobson shows that this equation, together with local conservation of energy (i.e., vanishing divergence of the energy–momentum tensor), can only be satisfied if

$$T_{ab} = \frac{c^4}{8\pi G} \left(R_{ab} - \left(\frac{1}{2} R + \Lambda \right) g_{ab} \right) \,, \tag{19}$$

where Λ is a constant of integration whose value is not determined by the problem. These are the full field equations of general relativity, including the cosmological constant Λ. This value of this constant remains undetermined, though.

The field equations are thus shown to be valid at horizons. Since it is possible, by choosing a suitable coordinate transformation, to position a horizon at any desired space-time event, the field equations must be valid over the whole of space-time.

Since it is possible to have a horizon at every event in space-time, there is the same maximum possible force (or power) at every event in nature. This maximum force (or power) is thus a constant of nature.

In other words, the field equations of general relativity are a direct consequence of the limited energy flow at horizons, which in turn is due to the existence of a maximum force or power. We can thus speak of the maximum force *principle*. Conversely, the field equations imply maximum force. Maximum force and general relativity are thus equivalent.

Deducing universal gravitation

Universal gravitation follows from the force limit in the case where both forces and speeds are much smaller than the maximum values. The first condition implies $\sqrt{4GMa} \ll c^2$, the second $v \ll c$ and $al \ll c^2$. Let us apply this to a specific case. Consider a satellite circling a central mass M at distance R with acceleration a. This

Challenge 6 e

system, with length $l = 2R$, has only one characteristic speed. Whenever this speed v is much smaller than c, v^2 must be proportional both to the squared speed calculated by $al = 2aR$ and to the squared speed calculated from $\sqrt{4GMa}$. Taken together, these two conditions imply that $a = fGM/R^2$, where f is a numerical factor. A quick check, for example using the observed escape velocity values, shows that $f = 1$. Forces and speeds much smaller than the limit values thus imply that the inverse square law of gravity holds. In other words, nature's limit on force implies the universal law of gravity. Challenge 7 e

THE SIZE OF PHYSICAL SYSTEMS IN GENERAL RELATIVITY

General relativity, like the other theories of modern physics, implies a limit on the *size* l of systems. There is a limit to the amount of matter that can be concentrated into a small volume:

$$l \geqslant \frac{4Gm}{c^2} \, . \qquad (20)$$

The size limit is only realized for *black holes*, those well-known systems which swallow everything that is thrown into them. It is fully equivalent to the force limit. All *composite* systems in nature comply with the lower size limit. Whether elementary particles fulfil or even match this limit remains open at this point. More about this issue below. Page 265

Ref. 20 General relativity also implies an 'indeterminacy relation' for the size l of systems:

$$\frac{\Delta E}{\Delta l} \leqslant \frac{c^4}{4G} \, . \qquad (21)$$

Experimental data are available only for composite systems; all known systems comply with it. For example, the latest measurements for the Sun give $T_{\odot} = GM_{\odot}/c^3 = 4.925\,490\,947\,\mu s$; the error in E is thus much smaller than the (scaled) error in its radius. This indeterminacy relation is not as well known as that from quantum theory. In fact, tests of it – for example with binary pulsars – may distinguish general relativity from competing theories. We cannot yet say whether this inequality also holds for elementary particles. Ref. 21

A MECHANICAL ANALOGY FOR THE MAXIMUM FORCE

The maximum force is central to the theory of general relativity. Indeed, its value (adorned with a factor 2π) appears in the field equations. Its importance becomes clearer when we return to our old image of space-time as a deformable mattress. Like any material body, a mattress is described by a material constant that relates the deformation values to the values of applied energy. Similarly, a mattress, like any material, is described by the maximum stress it can bear before it breaks. These two values describe all materials, from crystals to mattresses. In fact, for perfect crystals (without dislocations), these two material constants are the same.

Empty space somehow behaves like a perfect crystal, or a perfect mattress: it has a deformation-energy constant that is equal to the maximum force that can be applied to it. The constant of gravitation thus determines the elasticity of space-time. Now, materials are not homogeneous: crystals are made up of atoms, and mattresses are made up of foam bubbles. What is the corresponding structure of space-time? This is a central question in the rest of our adventure. One thing is sure: unlike crystals, vacuum has no preferred directions.

We now take a first step towards answering the question of the structure of space-time and particles by putting together all the limits found so far.

PLANCK LIMITS FOR ALL PHYSICAL OBSERVABLES

The existence of a maximum force in nature is equivalent to general relativity. As a result, a large part of modern physics can be summarized in four simple and fundamental statements on motion:

$$\begin{aligned} &\text{quantum theory on action:} && W \geqslant \hbar \\ &\text{thermodynamics on entropy:} && S \geqslant k \\ &\text{special relativity on speed:} && v \leqslant c \\ &\text{general relativity on force:} && F \leqslant \frac{c^4}{4G} \,. \end{aligned} \tag{22}$$

These limits are valid for all physical systems, whether composite or elementary, and for all observers. Note that the limit quantities of special relativity, thermodynamics, quantum theory and general relativity can also be seen as the right-hand sides of the respective indeterminacy relations. Indeed, the set (4, 7, 9, 21) of indeterminacy relations is fully equivalent to the four limit statements (22). Challenge 8 e

By combining the three fundamental limits, we can obtain limits on a number of physical observables. The following limits are valid generally, for both composite and elementary systems:

$$\text{time interval:} \quad t \geqslant \sqrt{\frac{4G\hbar}{c^5}} = 1.1 \cdot 10^{-43}\,\text{s} \tag{23}$$

$$\text{time-distance product:} \quad td \geqslant \frac{4G\hbar}{c^4} = 3.5 \cdot 10^{-78}\,\text{ms} \tag{24}$$

$$\text{acceleration:} \quad a \leqslant \sqrt{\frac{c^7}{4G\hbar}} = 2.8 \cdot 10^{51}\,\text{m/s}^2 \tag{25}$$

$$\text{angular frequency:} \quad \omega \leqslant 2\pi\sqrt{\frac{c^5}{2G\hbar}} = 5.8 \cdot 10^{43}\,/\text{s} \tag{26}$$

Adding the knowledge that space and time can mix, we get

$$\text{distance:} \quad d \geqslant \sqrt{\frac{4G\hbar}{c^3}} = 3.2 \cdot 10^{-35}\,\text{m} \tag{27}$$

$$\text{area:} \quad A \geqslant \frac{4G\hbar}{c^3} = 1.0 \cdot 10^{-69}\,\text{m}^2 \tag{28}$$

$$\text{volume:} \qquad V \geqslant \left(\frac{4G\hbar}{c^3}\right)^{3/2} = 3.4 \cdot 10^{-104}\,\mathrm{m}^3 \qquad (29)$$

$$\text{curvature:} \qquad K \leqslant \frac{c^3}{4G\hbar} = 1.0 \cdot 10^{69}\,/\mathrm{m}^2 \qquad (30)$$

$$\text{mass density:} \qquad \rho \leqslant \frac{c^5}{16G^2\hbar} = 3.2 \cdot 10^{95}\,\mathrm{kg/m}^3 \qquad (31)$$

Of course, speed, action, angular momentum, entropy, power and force are also limited, as already stated. Up to a numerical factor, the limit for every physical observable corresponds to the Planck value. (The limit values are deduced from the commonly used Planck values simply by substituting $4G$ for G.) These limit values are the true *natural units* of nature. In fact, the ideal case would be to redefine the usual Planck values for all observables to these extremal values, by absorbing the numerical factor 4 into the respective definitions. In the following, we call the limit values the *corrected Planck units* and assume that the factors have been properly included. In other words, *every natural unit or (corrected) Planck unit is the limit value of the corresponding physical observable.*

Page 54 Most of these limit statements are found scattered throughout the research literature, though the numerical factors often differ. Each limit has attracted a string of publications. Ref. 22 The existence of a smallest measurable distance and time interval of the order of the Planck values is discussed in quantum gravity and string theory. The maximum curvature has been studied in quantum gravity; Ref. 23 it has important consequences for the 'beginning' of the universe, where it excludes any infinitely large or small observable. The maximum mass density appears regularly in discussions on the energy of the vacuum.

Note that the different dimensions of the four fundamental limits (22) in nature means that the four limits are *independent*. For example, quantum effects cannot be used to overcome the force limit; similarly, the power limit cannot be used to overcome the speed limit. There are thus four independent limits on motion in nature.

> Die Frage über die Gültigkeit der Voraussetzungen der Geometrie im Unendlichkleinen hängt zusammen mit der Frage nach dem innern Grunde der Massverhältnisse des Raumes. Bei dieser Frage, welche wohl noch zur Lehre vom Raume gerechnet werden darf, kommt die obige Bemerkung zur Anwendung, dass bei einer discreten Mannigfaltigkeit das Princip der Massverhältnisse schon in dem Begriffe dieser Mannigfaltigkeit enthalten ist, bei einer stetigen aber anders woher hinzukommen muss. Es muss also entweder das dem Raume zu Grunde liegende Wirkliche eine discrete Mannigfaltigkeit bilden, oder der Grund der Massverhältnisse ausserhalb, in darauf wirkenden bindenen Kräften, gesucht werden.[*]
>
> Bernhard Riemann, 1854, *Über die Hypothesen, welche der Geometrie zu Grunde liegen*.

Physics, mathematics and simplicity

The four limits of nature are astonishing. For many decades, a silent assumption has guided many – but not all – physicists: physics requires *difficult* mathematics.[**] The above summary shows the exact opposite. The essence of the important physical theories is extremely simple: special relativity, general relativity, thermodynamics and quantum theory are each based on a simple inequality.

The summary of a large part of physics withe inequalities is suggestive. The summary makes us dream that the description of the remaining parts of physics – gauge fields, elementary particles and the final theory – might be equally simple. This dream thus contrasts with the silent assumption that unification requires complex mathematics. Let us continue to explore where the dream of simplicity leads us to.

Limits to space, time and size

> Those are my principles, and if you don't like them ... well, I have others.
>
> Groucho Marx

The four fundamental limits of nature (22) result in a minimum distance and a minimum time interval. As the expressions for the limits shows, these minimum intervals arise directly from the *unification* of quantum theory and relativity: they do not appear if the

---

[*] 'The question of the validity of the hypotheses of geometry in the infinitely small is connected to the question of the foundation of the metric relations of space. To this question, which may still be regarded as belonging to the study of space, applies the remark made above; that in a discrete manifold the principles of its metric relations are given in the notion of this manifold, while in a continuous manifold, they must come from outside. Either therefore the reality which underlies space must form a discrete manifold, or the principles of its metric relations must be sought outside it, in binding forces which act upon it.'

[**] For example, for over thirty years, Albert Einstein searched with his legendary intensity for the final theory by exploring more and more complex equations. He did so even on his deathbed! Also most theoretical physicists in the year 2000 held the prejudice that unification requires complex mathematics. This prejudice is a consequence of over a century of flawed teaching of physics. The prejudice is also one of the reasons that the search for a final theory was not successful for so long.

theories are kept separate. In short, unification implies that there is a smallest length in nature. This result is important: the formulation of physics as a set of limit statements shows that *the continuum model of space and time is not completely correct*. Continuity and manifolds are only approximations, valid for large actions, low speeds and small forces. The reformulation of general relativity and quantum theory with limit statements makes this especially clear.

The existence of a force limit in nature implies that no physical system can be smaller than a Schwarzschild black hole of the same mass. In particular, *point particles do not exist*. The density limit makes the same point. In addition, elementary particles are predicted to be larger than the corrected Planck length. So far, this prediction has not been tested by observations, as the scales in question are so small that they are beyond experimental reach. Detecting the sizes of elementary particles – for example, with electric dipole measurements – would make it possible to check all limits directly. Page 54

Mass and energy limits

Mass plays a special role in all these arguments. The set of limits (22) does not make it possible to extract a limit statement on the mass of physical systems. To find one, we have to restrict our aim somewhat.

The Planck limits mentioned so far apply to *all* physical systems, whether composite or elementary. Other limits apply only to elementary systems. In quantum theory, the distance limit is a size limit only for *composite* systems. A particle is elementary if the system size l is smaller than any conceivable dimension:

$$\text{for elementary particles:} \quad l \leqslant \frac{\hbar}{mc} \,. \tag{32}$$

Using this limit, we find the well-known mass, energy and momentum limits, valid only for elementary particles:

$$\begin{aligned}
\text{for elementary particles:} \quad m &\leqslant \sqrt{\frac{\hbar c}{4G}} = 1.1 \cdot 10^{-8}\,\mathrm{kg} = 0.60 \cdot 10^{19}\,\mathrm{GeV/c^2} \\
\text{for elementary particles:} \quad E &\leqslant \sqrt{\frac{\hbar c^5}{4G}} = 9.8 \cdot 10^{8}\,\mathrm{J} = 0.60 \cdot 10^{19}\,\mathrm{GeV} \\
\text{for elementary particles:} \quad p &\leqslant \sqrt{\frac{\hbar c^3}{4G}} = 3.2\,\mathrm{kg\,m/s} = 0.60 \cdot 10^{19}\,\mathrm{GeV/c}
\end{aligned} \tag{33}$$

These elementary-particle limits, corresponding to the corrected Planck mass, energy and momentum, were discussed in 1968 by Andrei Sakharov, though with different numerical factors. They are regularly cited in elementary particle theory. All known measurements comply with them. Ref. 24

Virtual particles – a new definition

In fact, there are elementary particles that exceed all three limits that we have encountered so far. Nature does have particles that move faster than light, that show actions below the quantum of action, and that experience forces larger than the force limit.

Vol. II, page 65 Vol. IV, page 164 We know from special relativity that the virtual particles exchanged in collisions move faster than light. We know from quantum theory that virtual particle exchange implies actions below the minimum action. Virtual particles also imply an instantaneous change of momentum; they thus exceed the force limit. Thus *virtual particles* exceed all the limits that hold for *real* elementary particles.

Curiosities and fun challenges about Planck limits

The (corrected) Planck limits are statements about properties of nature. There is no way to measure values exceeding these limits, with any kind of experiment. Naturally, such a claim provokes the search for counter-examples and leads to many paradoxes.

* *

The minimum action may come as a surprise at first, because angular momentum and spin have the same unit as action; and nature contains particles with spin 0 or with spin $1/2\,\hbar$. A minimum action indeed implies a minimum angular momentum. However, the angular momentum in question is *total* angular momentum, including the orbital part with respect to the observer. The measured total angular momentum of a particle is never smaller than $\hbar$, even if the spin is smaller.

* *

A further way to deduce the minimum length using the limit statements which structure this ascent is the following. General relativity is based on a maximum force in nature, or alternatively, on a maximum mass change per time, whose value is given by $\mathrm{d}m/\mathrm{d}t = c^3/4G$. Quantum theory is based on a minimum action W in nature, given by $\hbar$. Since a distance d can be expressed as Vol. II, page 95

$$d^2 = \frac{W}{\mathrm{d}m/\mathrm{d}t} \,, \tag{34}$$

one sees directly that a minimum action and a maximum rate of change of mass imply a minimum distance. In other words, quantum theory and general relativity force us to conclude that *in nature there is a minimum distance.* In other words, *at Planck scales the term 'point in space' has no theoretical or experimental basis.*

* *

With the single-particle limits, the entropy limit leads to an upper limit for temperature:

$$T \leqslant \sqrt{\frac{\hbar c^5}{4Gk^2}} = 0.71 \cdot 10^{32}\,\mathrm{K} \,. \tag{35}$$

This corresponds to the temperature at which the energy per degree of freedom is given by the (corrected) Planck energy $\sqrt{\hbar c^5/4G}$. A more realistic value would have to take account of the number of degrees of freedom of a particle at Planck energy. This would change the numerical factor. However, no system that is even near this temperature value has been studied yet. Only horizons are expected to realize the temperature limit, but nobody has managed to explore them yet.

* *

How can the maximum force be determined by gravity alone, which is the weakest interaction? It turns out that in situations near the maximum force, the other interactions are negligible. This is the reason why gravity must be included in a unified description of nature.

* *

Challenge 9 ny

At first sight, it seems that electric charge can be used in such a way that the acceleration of a charged body towards a charged black hole is increased to a value, when multiplied with the mass, that exceeds the force limit. However, the changes in the horizon for charged black holes prevent this.

* *

The gravitational attraction between two masses never yields force values high enough to exceed the force limit. Why? First of all, masses m and M cannot come closer together than the sum of their horizon radii. Using $F = GmM/r^2$ with the distance r given by the (naive) sum of the two black hole radii as $r = 2G(M+m)/c^2$, we get

$$F \leqslant \frac{c^4}{4G} \frac{Mm}{(M+m)^2} , \tag{36}$$

which is never larger than the force limit. Thus even two attracting black holes cannot exceed the force limit – in the inverse-square approximation of universal gravity. In short, the minimum size of masses means that the maximum force cannot be exceeded.

* *

Ref. 25

It is well known that gravity bends space. Therefore, if they are to be fully convincing, our calculation needs to be repeated taking into account the curvature of space. The simplest way is to study the force generated by a black hole on a test mass hanging from a wire that is lowered towards a black hole horizon. For an *unrealistic point mass*, the force would diverge at the horizon. Indeed, for a point mass m lowered towards a black hole of mass M at (conventionally defined radial) distance d, the force would be

$$F = \frac{GMm}{d^2 \sqrt{1 - \frac{2GM}{dc^2}}} . \tag{37}$$

This diverges at $d = 0$, the location of the horizon. However, even a test mass cannot be smaller than its own gravitational radius. If we want to reach the horizon with a *realistic* test mass, we need to choose a small test mass m: only a small mass can get near the horizon. For vanishingly small masses, however, the resulting force tends to zero. Indeed, letting the distance tend to the smallest possible value by letting $d = 2G(m + M)/c^2 \rightarrow 2GM/c^2$ requires $m \rightarrow 0$, which makes the force $F(m, d)$ vanish. If on the other hand, we remain away from the horizon and look for the maximum force by using a mass as large as can possibly fit into the available distance (the calculation is straightforward), then again the force limit is never exceeded. In other words, for *realistic* test masses, expression (37) is *never* larger than $c^4/4G$. Taking into account the minimal size of test masses, we thus see that the maximum force is never exceeded in gravitational systems.

* *

An absolute power limit implies a limit on the energy that can be transported per unit time through any imaginable surface. At first sight, it may seem that the combined power emitted by two radiation sources that each emit 3/4 of the maximum value should give 3/2 times the maximum value. However, the combination forms a black hole, or at least prevents part of the radiation from being emitted by swallowing it between the two sources. Challenge 10 e

* *

One possible system that actually achieves the power limit is the final stage of black hole evaporation. But even in this case, the power limit is not exceeded. Challenge 11 e

* *

The maximum force limit states that the stress-energy tensor, when integrated over any physical surface, does not exceed the limit value. Ref. 16 No such integral, over any physical surface, of any tensor component in any coordinate system, can exceed the force limit, provided that it is measured by a nearby observer or a test body with a realistic proper size. The maximum force limit thus applies to any component of any force vector, as well as to its magnitude. It applies to gravitational, electromagnetic, and nuclear forces; and it applies to all realistic observers. It is not important whether the forces are real or fictitious; nor whether we are discussing the 3-forces of Galilean physics or the 4-forces of special relativity. Indeed, the force limit applied to the zeroth component of the 4-force is the power limit.

* *

In terms of mass flows, the power limit implies that flow of water through a tube is limited in throughput. The resulting limit $\mathrm{d}m/\mathrm{d}t \leqslant c^3/4G$ for the change of mass with time seems to be unrecorded in the literature of the twentieth century.

* *

The force limit cannot be overcome with Lorentz boosts. One might think that a boost can be chosen in such a way that a 3-force value F in one frame is transformed into any desired value F' in another, boosted frame. This thought turns out to be wrong. In Vol. II, page 74 relativity, 3-force cannot be increased beyond all bounds using boosts. In all reference frames, the measured 3-force can never exceed the proper force, i.e., the 3-force value measured in the comoving frame.

* *

The power limit is of interest if applied to the universe as a whole. Indeed, it can be used to explain Olbers' paradox: the sky is dark at night because the combined luminosity of all light sources in the universe cannot be brighter than the maximum value.

* *

Page 32 The force limit and its solid state analogy might be seen to suggest that the appearance of matter might be nature's way of preventing space-time from ripping apart. Does this analogy make sense? Challenge 12 ny

* *

Ref. 26 In fact, the connection between minimum length and gravity is not new. Already in 1967, Andrei Sakharov pointed out that a minimum length implies gravity. He showed that regularizing quantum field theory on curved space with a cut-off will induce counter-terms that include to lowest order the cosmological constant and then the Einstein–Hilbert action.

* *

The existence of a smallest length – and a corresponding shortest time interval – implies that no surface is *physical* if any part of it requires a localization in space-time to scales below the minimum length. (In addition, a physical surface must not cross any horizon.) Only by insisting on this condition can we eliminate unphysical examples that contravene the force and power limits. For example, this condition was overlooked in Bousso's early discussion of Bekenstein's entropy bound – though not in his more recent ones. Ref. 27

* *

Our discussion of limits can be extended to include electromagnetism. Using the (low-energy) electromagnetic coupling constant α, we get the following limits for physical systems interacting electromagnetically:

$$\text{electric charge:} \qquad q \geqslant \sqrt{4\pi\varepsilon_0 \alpha c\hbar} = e = 0.16\,\text{aC} \tag{38}$$

$$\text{electric field:} \qquad E \leqslant \sqrt{\frac{c^7}{64\pi\varepsilon_0 \alpha \hbar G^2}} = \frac{c^4}{4Ge} = 1.9 \cdot 10^{62}\ \text{V/m} \tag{39}$$

$$\text{magnetic field:} \qquad B \leqslant \sqrt{\frac{c^5}{64\pi\varepsilon_0 \alpha \hbar G^2}} = \frac{c^3}{4Ge} = 6.3 \cdot 10^{53}\ \text{T} \tag{40}$$

$$\text{voltage:} \qquad U \leqslant \sqrt{\frac{c^4}{16\pi\varepsilon_0 \alpha G}} = \frac{1}{e}\sqrt{\frac{\hbar c^5}{4G}} = 6.1 \cdot 10^{27}\ \text{V} \tag{41}$$

$$\text{inductance:} \qquad L \geqslant \frac{1}{4\pi\varepsilon_0 \alpha}\sqrt{\frac{4G\hbar}{c^7}} = \frac{1}{e^2}\sqrt{\frac{4G\hbar^3}{c^5}} = 4.4 \cdot 10^{-40}\ \text{H} \tag{42}$$

With the additional assumption that in nature at most one particle can occupy one Planck volume, we get

$$\text{charge density:} \quad \rho_e \leqslant \sqrt{\frac{\pi\varepsilon_0\alpha}{16G^3}}\,\frac{c^5}{\hbar} = e\sqrt{\frac{c^9}{64G^3\hbar^3}} = 4.7\cdot 10^{84}\,\mathrm{C/m^3} \tag{43}$$

$$\text{capacitance:} \quad C \geqslant 4\pi\varepsilon_0\alpha\sqrt{\frac{4G\hbar}{c^3}} = e^2\sqrt{\frac{4G}{c^5\hbar}} = 2.6\cdot 10^{-47}\,\mathrm{F} \tag{44}$$

For the case of a single conduction channel, we get

$$\text{electric resistance:} \quad R \geqslant \frac{1}{4\pi\varepsilon_0\alpha c} = \frac{\hbar}{e^2} = 4.1\,\mathrm{k\Omega} \tag{45}$$

$$\text{electric conductivity:} \quad G \leqslant 4\pi\varepsilon_0\alpha c = \frac{e^2}{\hbar} = 0.24\,\mathrm{mS} \tag{46}$$

$$\text{electric current:} \quad I \leqslant \sqrt{\frac{\pi\varepsilon_0\alpha c^6}{G}} = e\sqrt{\frac{c^5}{4\hbar G}} = 1.5\cdot 10^{24}\,\mathrm{A} \tag{47}$$

The magnetic field limit is significant in the study of extreme stars and black holes. The maximum electric field plays a role in the theory of gamma ray bursters. For current, conductivity and resistance in single channels, the limits and their effects were studied extensively in the 1980s and 1990s by researchers who will probably win a Nobel Prize in the not too distant future.

Ref. 28

Ref. 29

The observation of quarks and of collective excitations in semiconductors with charge $e/3$ does not necessarily invalidate the charge limit for physical systems. In neither case is there is a physical system – defined as localized mass–energy interacting incoherently with the environment – with charge $e/3$.

* *

The general fact that to every limit value in nature there is a corresponding indeterminacy relation is valid in particular for electricity. Indeed, there is an indeterminacy relation for capacitors, of the form

$$\Delta C\ \Delta U \geqslant e \tag{48}$$

where e is the positron charge, C capacity and U potential difference; there is also an indeterminacy relation between electric current I and time t

$$\Delta I\ \Delta t \geqslant e\ . \tag{49}$$

Ref. 30 Both these relations may be found in the literature.

COSMOLOGICAL LIMITS FOR ALL PHYSICAL OBSERVABLES

In our quest to understand motion, we have focused our attention on the four fundamental limitations to which it is subject. Special relativity posits a limit to speed, namely the speed of light c. General relativity limits force and power respectively by $c^4/4G$ and $c^5/4G$, and quantum theory introduces a smallest value $\hbar$ for action. Nature imposes the limit k on entropy. If we include the limit e on electric charge changes, these limits induce extremal values for *all* physical observables, given by the corresponding (corrected) Planck values. Vol. II, page 95

A question arises: does nature also impose limits on physical observables at the opposite end of the measurement scale? For example, there is a highest force and a highest power in nature. Is there also a lowest force and a lowest power? Is there also a lowest speed?

We will show that there are indeed such limits, for all observables. We give the general method to generate such bounds, and explore several examples. This exploration will take us on an interesting survey of modern physics; we start by deducing system-dependent limits and then go on to the cosmological limits.

Size and energy dependence

While looking for additional limits in nature, we note a fundamental fact. Any upper limit for angular momentum, and any lower limit for power, must be *system-dependent*. Such limits will not be absolute, but will depend on properties of the system. Now, a physical system is a part of nature characterized by a boundary and its content.* Thus the simplest properties shared by all systems are their size (characterized in the following by the diameter) L and their energy E. With these characteristics we can deduce system-dependent limits for every physical observable. The general method is straightforward: we take the known inequalities for speed, action, power, charge and entropy, and then extract a limit for any observable, by inserting the length and energy as required. We then have to select the strictest of the limits we find.

Angular momentum and action

It only takes a moment to check that the ratio of angular momentum D to energy E times length L has the dimensions of inverse speed. Since speeds are limited by the speed of light, we get Challenge 13 e

$$D_{\text{system}} \leqslant \frac{1}{c} LE . \qquad (50)$$

Indeed, in nature there do not seem to be any exceptions to this limit on angular momentum. In no known system, from atoms to molecules, from ice skaters to galaxies, does the angular momentum exceed this value. Even the most violently rotating objects, the so-called extremal black holes, are limited in angular momentum by $D \leqslant LE/c$. (Actually, this limit is correct for black holes only if the energy is taken as the irreducible Ref. 31

---

* Quantum theory refines this definition: a physical system is a part of nature that in addition interacts *incoherently* with its environment. In the following discussion we will assume that this condition is satisfied. Vol. IV, page 127

mass times c^2; if the usual mass is used, the limit is too large by a factor of 4.) The limit deduced from general relativity, given by $D \leqslant L^2 c^3/4G$, is not stricter than the one just given. No system-dependent lower limit for angular momentum can be deduced.

The maximum value for angular momentum is also interesting when it is seen as an action limit. Action is the time integral of the difference between kinetic and potential energy. Since nature always seeks to minimize the action W, it seems strange to search for systems that maximize it. You might check for yourself that the action limit

$$W \leqslant LE/c \tag{51}$$

Challenge 14 ny is not exceeded in any physical process.

Speed

Speed times mass times length is an action. Since action values in nature are limited from below by $\hbar$, we get a limit for the speed of a system:

$$v_{\text{system}} \geqslant \hbar c^2 \frac{1}{LE} \; . \tag{52}$$

This is not a new result; it is just a form of the indeterminacy relation of quantum theory. It gives a minimum speed for any system of energy E and diameter L. Even the extremely slow radius change of a black hole by evaporation just realizes this minimal speed. Continuing with the same method, we also find that the limit deduced from general relativity, $v \leqslant (c^2/4G)(L/E)$, gives no new information. Therefore, no *system-dependent* upper speed limit exists. Challenge 15 ny Challenge 16 e

Incidentally, the limits are not unique. Other limits can be found in a systematic way. Upper limits can be multiplied, for example, by factors of $(L/E)(c^4/4G)$ or $(LE)(2/\hbar c)$, yielding less strict upper limits. A similar rule can be given for lower limits.*

Force, power and luminosity

We have seen that force and power are central to general relativity. The force exerted by a system is the flow of momentum out of the system; emitted power is the flow of energy out of the system. Thanks to the connection $W = FLT$ between action W, force F, distance L and time T, we can deduce

$$F_{\text{system}} \geqslant \frac{\hbar}{2c} \frac{1}{T^2} \; . \tag{53}$$

Experiments do not reach this limit. The smallest forces measured in nature are those in atomic force microscopes, where values as small as 1 aN are observed. But even these values are above the lower force limit.

---

* The strictest upper limits are those with the smallest exponent for length, and the strictest lower limits are those with the largest exponent of length.

The power P emitted by a system of size L and mass M is limited by

$$c^3 \frac{M}{L} \geqslant P_{\text{system}} \geqslant 2\hbar G \frac{M}{L^3} \,. \tag{54}$$

The limit on the left is the upper limit for any engine or lamp, as deduced from relativity; not even the universe exceeds it. The limit on the right is the minimum power emitted by any system through quantum gravity effects. Indeed, no physical system is completely tight. Even black holes, the systems with the best ability to keep components inside their enclosure, radiate. The power radiated by black holes should just meet this limit, provided the length L is taken to be the circumference of the black hole. Thus the claim of the quantum gravity limit is that the power emitted by a black hole is the smallest power that is emitted by any composite system of the same surface gravity. (However, the numerical factors in the black hole power appearing in the literature are not yet consistent.)

THE STRANGE CHARM OF THE ENTROPY BOUND

Ref. 32 In 1973, Bekenstein discovered a famous limit that connects the entropy S of a physical system with its size and mass. No system has a larger entropy than one bounded by a horizon. The larger the horizon surface, the larger the entropy. We write

$$\frac{S}{S_{\text{limit}}} \leqslant \frac{A}{A_{\text{limit}}} \tag{55}$$

which gives

$$S \leqslant \frac{kc^3}{4G\hbar} A \,, \tag{56}$$

where A is the surface of the system. Equality is realized only for black holes. The old question of the origin of the factor 4 in the entropy of black holes is thus answered here: it is due to the factor 4 in the force or power bound in nature. Time will tell whether this explanation will be generally accepted. Stay tuned.

We can also derive a more general relation by using a mysterious assumption, which we will discuss afterwards. We assume that the limits for vacuum are opposite to those for matter. We can then write $c^2/4G \leqslant M/L$ for the vacuum. Using

$$\frac{S}{S_{\text{c.Planck}}} \leqslant \frac{M}{M_{\text{c.Planck}}} \frac{A}{A_{\text{c.Planck}}} \frac{L_{\text{c.Planck}}}{L} \tag{57}$$

we get

$$S \leqslant \frac{\pi k c}{\hbar} ML = \frac{2\pi k c}{\hbar} MR \,. \tag{58}$$

This is called *Bekenstein's entropy bound*. No exception has ever been found or constructed, despite many attempts. Again, the limit value itself is only realized for black holes. Ref. 27

We need to explain the strange assumption used above. We are investigating the entropy of a horizon. Horizons are not matter, but limits to empty space. The entropy of horizons is due to the large number of virtual particles found at them. In order to deduce the maximum entropy of expression (57) one therefore has to use the properties of the vacuum. In other words, *either* we use a mass-to-length ratio for vacuum *above* the Planck limit, *or* we use the Planck entropy as the *maximum* value for vacuum.

Other, equivalent limits for entropy can be found if other variables are introduced. For example, since the ratio of the shear viscosity η to the volume density of entropy (times k) has the dimensions of action, we can directly write Ref. 34

$$S \leqslant \frac{k}{\hbar}\,\eta V \ . \qquad (59)$$

Again, equality is only attained in the case of black holes. In time, no doubt, the list of similar bounds will grow longer.

Is there also a smallest, system-dependent entropy? So far, there does not seem to be a system-dependent minimum value for entropy: the present approach gives no expression that is larger than k. Challenge 17 ny

The establishment of the entropy limit is an important step towards making our description of motion consistent. If space-time can move, as general relativity maintains, it also has an entropy. How could entropy be limited if space-time were continuous? Clearly, because of the existence of a minimum distance and minimum time in nature, space-time cannot be continuous, but must have a finite number of degrees of freedom, and thus a finite entropy.

CURIOSITIES AND FUN CHALLENGES ABOUT SYSTEM-DEPENDENT LIMITS TO OBSERVABLES

Also the system-dependent limit values for all physical observables, like the Planck values, yield a plethora of interesting questions. We study a few examples.

* *

The content of a system is characterized not only by its mass and charge, but also by its strangeness, isospin, colour charge, charge and parity. Can you deduce the limits for these quantities? Challenge 18 r

* *

In our discussion of black hole limits, we silently assumed that they interact, like any thermal system, in an incoherent way with the environment. Which of the results of this section change when this condition is dropped, and how? Which limits can be overcome? Challenge 19 s

* *

Challenge 20 e Can you find a general method to deduce all limits of observables?

* *

The Bekenstein's entropy limit leads to some interesting speculations. Let us speculate that the universe itself, being surrounded by a horizon, meets the Bekenstein bound. The entropy bound gives a bound to all degrees of freedom inside a system: it tells us that the number $N_{\text{d.o.f.}}$ of degrees of freedom in the universe is roughly

Challenge 21 e

$$N_{\text{d.o.f.}} \approx 10^{132} \, . \tag{60}$$

Compare this with the number $N_{\text{Pl. vol.}}$ of Planck volumes in the universe

$$N_{\text{Pl. vol.}} \approx 10^{183} \tag{61}$$

and with the number $N_{\text{part.}}$ of particles in the universe

$$N_{\text{part.}} \approx 10^{91} \, . \tag{62}$$

We see that particles are only a tiny fraction of what moves around. Most motion must be movement of space-time. At the same time, space-time moves much less than might be naively expected. To find out how all this happens is the challenge of the unified description of motion.

* *

A lower limit for the temperature of a thermal system can be found using the idea that the number of degrees of freedom of a system is limited by its surface, or more precisely, by the ratio between the surface and the Planck surface. We get the limit

$$T \geqslant \frac{4G\hbar}{\pi k c} \frac{M}{L^2} \, . \tag{63}$$

Alternatively, using the method given above, we can use the limit on the thermal energy $kT/2 \geqslant \hbar c/2\pi L$ (the thermal wavelength must be smaller than the size of the system) together with the limit on mass $c^2/4G \geqslant M/L$, and deduce the same result.

We have met the temperature limit already: when the system is a black hole, the limit yields the temperature of the emitted radiation. In other words, the temperature of black holes is the lower limit for all physical systems for which a temperature can be defined, provided they share the *same boundary gravity*. The latter condition makes sense: boundary gravity is accessible from the outside and describes the full physical system, since it depends on both its boundary and its content. So far, no exception to this claim is known. All systems from everyday life comply with it, as do all stars. Even the coldest known systems in the universe, namely Bose–Einstein condensates and other cold matter produced in laboratories, are much hotter than the limit, and thus much hotter than black holes of the same surface gravity. (We saw earlier that a consistent Lorentz transformation for temperature is not possible; so the minimum temperature limit is only valid for an observer at the same gravitational potential as the system under consideration and stationary relative to it.)

Challenge 22 s

Vol. II, page 57

There seems to be no consistent way to define an upper limit for a system-dependent temperature. Limits for other thermodynamic quantities can be found, but we will not

Challenge 23 ny

discuss them here.

* *

When electromagnetism plays a role in a system, the system also needs to be characterized by a charge Q. Our method then gives the following lower limit for the electric field E:

$$E \geqslant 4Ge \frac{M^2}{Q^2 L^2} . \tag{64}$$

We write the field limit in terms of the elementary charge e, though it would be more appropriate to write it using the fine structure constant via $e = \sqrt{4\pi\varepsilon_0 \alpha \hbar c}$. In observations, the electric field limit has never been exceeded. Can you show whether it is attained by maximally charged black holes? Challenge 24 ny

For the magnetic field we get

$$B \geqslant \frac{4Ge}{c} \frac{M^2}{Q^2 L^2} . \tag{65}$$

Again, this limit is satisfied by all known systems in nature.

Similar limits can be found for the other electromagnetic observables. In fact, several of the limits given earlier are modified when electric charge is included. Can you show how the size limit changes when electric charge is taken into account? There is an entire research field dedicated to the deduction of the most general limits valid in nature. Challenge 25 ny

Cosmology in one statement

We now continue our exploration of limits to the largest systems possible. In order to do that, we have a simple look at cosmology.

Cosmology results from the equations of general relativity when the cosmological constant is included. Cosmology can thus be summarized by any sufficiently general statement that includes the cosmological constant Λ. The simplest statement can be deduced from the observation that the present distance R_0 of the night sky horizon is about $R_0 \approx 1/\sqrt{\Lambda}$. From this we can summarize cosmology by the inequality

$$l \lessapprox \frac{1}{\sqrt{\Lambda}} . \tag{66}$$

This statement contains all of cosmology; at present, the precise numerical factor is not of importance. This statement must be added as a fifth statement on physics to the four fundamental Planck limits.

Challenge 26 ny By the way, can you show that the cosmological constant is observer-invariant?

The cosmological limits to observables

From the system-dependent limits for speed, action, force and entropy we can deduce system-dependent limits for all other physical observables. In addition, we note that the system-dependent limits can (usually) be applied to the universe as a whole; we only need to insert the size and energy content of the universe. Usually, we can do this through a limit process, even though the universe itself is not a physical system. In this way, we get an absolute limit for every physical observable that contains the cosmological constant Λ and that is on the opposite end of the Planck limit for that observable. We can call these limits the *cosmological limits*.

The simplest cosmological limit is the upper limit to length in the universe. Since the cosmological length limit also implies a maximum possible Compton wavelength, we get a minimum particle mass and energy. We also get an cosmological lower limit on luminosity, etc.

For *single particles*, we find the absolute lower speed limit, the cosmological speed limit, given by

$$v_{\text{particle}} \geqslant \frac{\sqrt{4G\hbar/c}}{L_{\text{universe}}} = L_{\text{corr. Planck}} \sqrt{\Lambda}\, c \approx 7 \cdot 10^{-53}\,\text{m/s} . \qquad (67)$$

It has never been reached or approached by any observation.

The negative energy volume density $-\Lambda c^4/4\pi G$ introduced by the positive cosmological constant Λ corresponds to a negative pressure (both quantities have the same dimensions). When multiplied by the minimum area it yields a force value

$$F = \frac{\Lambda\hbar c}{2\pi} = 4.8 \cdot 10^{-79}\,\text{N} . \qquad (68)$$

Apart from the numerical factor, this is the *cosmological force limit*, the smallest possible force in nature. This is also the gravitational force between two corrected Planck masses located at the cosmological distance $\sqrt{\pi/4\Lambda}$.

As a note, we are led the fascinating conjecture that the full theory of general relativity, including the cosmological constant, is defined by the combination of a maximum and a minimum force in nature.

Another note concerns the importance of black hole limits for the universe itself. The observed average mass density of the universe is not far from the corresponding black hole limit. The black hole lifetime limit might thus provide an upper limit for the full lifetime of the universe. However, the age of the universe is far from that limit by a large factor. In fact, since the universe's size and age are increasing, the lifetime limit is pushed further into the future with every second that passes. The universe evolves so as to escape its own decay. Challenge 27 e

In summary, *nature provides two limits for each observable: a Planck limit and a cosmological limit.* The existence of two limits for each observable, a lower and an upper one, has important consequences that we will explore now.

Limits to measurement precision and their challenge to thought

We now know that in nature, every physical measurement has a lower and an upper bound. One of the bounds is cosmological, the other is given by the (corrected) Planck unit. As a consequence, for every observable, the smallest relative measurement error that is possible in nature is the ratio between the Planck limit and the cosmological limit.

In particular, we have to conclude that *all measurements are limited in precision.*

All limits, those to observables and those to measurement precision, only appear when quantum theory and gravity are brought together. But the existence of these limits, and in particular the existence of limits to measurement precision, forces us to abandon some cherished assumptions.

No real numbers

Because of the fundamental limits to measurement precision, *the measured values of physical observables do not require the full set of real numbers.* In fact, limited precision implies that observables cannot be described by the real numbers! This staggering result appears whenever quantum theory and gravity are brought together. But there is more.

Vacuum and mass: two sides of the same coin

There is a limit to the precision of length measurements in nature. This limit is valid both for length measurements of empty space and for length measurements of matter (or radiation). Now let us recall what we do when we measure the length of a table with a ruler. To find the ends of the table, we must be able to distinguish the table from the surrounding air. In more precise terms, we must be able to distinguish matter from vacuum.

When we want high measurement precision, we need to approach Planck scales. But at Planck scales, the measurement values and the measurement errors are of the same size. In short, at Planck scale, the intrinsic measurement limitations of nature imply that we cannot say whether we are measuring vacuum or matter. We will check this conclusion in detail later on. Page 54

In fact, we can pick any other observable that distinguishes vacuum from matter – for example, colour, mass, size, charge, speed or angular momentum – and we have the same problem: at Planck scales, the limits to observables lead to limits to measurement precision, and therefore, at Planck scales it is impossible to distinguish between matter and vacuum. At Planck scales, we cannot tell whether a box is full or empty.

To state the conclusion in the sharpest possible terms: *vacuum and matter do not differ at Planck scales*. This counter-intuitive result is one of the charms of the search for a final, unified theory. It has inspired many researchers in the field and some have written best-sellers about it. Brian Greene was particularly successful in presenting this side of quantum geometry to the wider public. Ref. 35

Limited measurement precision also implies that at the Planck energy it is impossible to speak about points, instants, events or dimensionality. Similarly, at the Planck length it is impossible to distinguish between positive and negative time values: so particles and antiparticles are not clearly distinguished at Planck scales. All these conclusions are so far-reaching that we must check them in more detail. We will do this shortly.

Measurement precision and the existence of sets

In physics, it is generally assumed that nature is a *set* of components. These components, called *elements* by mathematicians, are assumed to be *separable* from each other. This tacit assumption is introduced in three main situations: it is assumed that matter consists of separable particles, that space-time consists of separable events or points, and that the set of states consists of separable initial conditions. Until the year 2000, physics has built the whole of its description of nature on the concept of a set.

A fundamental limit to measurement precision implies that nature is *not* a set of such separable elements. Precision limits imply that physical entities can be distinguished only *approximately*. The approximate distinction is only possible at energies much lower than the Planck energy $\sqrt{\hbar c^5/4G}$. As humans, we do live at such small energies, and we can safely make the approximation. Indeed, the approximation is excellent in practice; we do not notice any error. But at Planck energy, distinction and separation is impossible in principle. In particular, at the cosmic horizon, at the big bang, and at Planck scales, any precise distinction between two events, two points or two particles becomes impossible.

Page 54

Another way to reach this result is the following. Separation of two entities requires *different measurement results* – for example, different positions, different masses or different velocities. Whatever observable is chosen, at the Planck energy the distinction becomes impossible because of the large measurements errors. Only at everyday energies is a distinction approximately possible. Any distinction between two physical systems – for example, between a toothpick and a mountain – is possible only *approximately*: at Planck scales, a boundary cannot be drawn.

Page 54

A third argument is the following. In order to *count* any entities in nature – a set of particles, a discrete set of points, or any other discrete set of physical observables – the entities have to be separable. But the inevitable measurement errors contradict separability. Thus at the Planck energy it is impossible to count physical objects with precision. *Nature has no parts.*

Page 54

In summary, at Planck scales, perfect separation is impossible in principle. We cannot distinguish observations. *At Planck scales it is impossible to split nature into separate entities.* In nature, elements of sets cannot be defined. Neither discrete nor continuous sets can be constructed. *Nature does not contain sets or elements.*

Page 54

Since sets and elements are only approximations, the concept of a 'set', which assumes separable elements, is too specialized to describe nature. Nature cannot be described at Planck scales – i.e., with full precision – if any of the concepts used for its description presupposes sets. However, all concepts used in the past 25 centuries to describe nature – space, time, particles, phase space, observables, wave functions, Hilbert space, Fock space, Riemannian space, particle space, loop space or moduli space – are based on sets. They must all be abandoned at Planck energy.

In short, nature is one and has no parts. *No correct mathematical model of nature can be based on sets.* But none of the approaches used in the twentieth century in theoretical physics has abandoned sets. This requirement is thus very powerful; indeed, it will guide us in the search for the unification of relativity and quantum theory. The requirement will even solve Hilbert's sixth problem.

Page 103

> “Es ist fast unmöglich, die Fackel der Wahrheit durch ein Gedränge zu tragen, ohne jemandem den Bart zu sengen.*”
>
> Georg Christoph Lichtenberg (1742–1799)

---

* 'It is almost impossible to carry the torch of truth through a crowd without scorching somebody's beard.'

SUMMARY ON LIMITS IN NATURE

If we exclude gauge interactions, we can summarize the rest of physics in a few limit statements. The speed limit is equivalent to special relativity, the force limit to general relativity, the action limit to quantum theory, the entropy limit to thermodynamics and the distance limit to cosmology. These limits are observer-invariant.

The invariant limits suggest interesting Gedanken experiments, none of which leads to their violation. The invariant limits imply that in nature every physical observable is bound on one end by the corresponding (corrected) Planck unit and on the other end by a cosmological limit. Every observable in nature has an upper and lower limit value.

The existence of lower and upper limit values to all observables implies that measurement precision is limited. As a consequence, matter and vacuum are indistinguishable, the description of space-time as a continuous manifold is not correct, and nature can be described by sets only approximately.

Since the most precise physical theories known, quantum theory and general relativity, can be reduced to limit statements, there is a good chance that the final, unified theory of physics will require an equally simple description. Nature's limits thus suggest that the mathematics of the final, unified theory might be simple. In addition, nature's limits imply that Planck units are the key to the final theory.

Page 18

At this point of our adventure, many questions are still open. Answering any of the open issues of the millennium list still seems out of reach. But this impression is too pessimistic. Our discussion implies that we only need to find a description of nature *without sets*. And a natural way to avoid the use of sets is a description of empty space, radiation and matter as being made of *common* constituents. But before we explore this option, we check the conclusions of this chapter in another way. In particular, as a help to more conservative physicists, we check all conclusions we found so far *without* making use of the maximum force principle.

CHAPTER 3

GENERAL RELATIVITY VERSUS QUANTUM THEORY

> “Man muß die Denkgewohnheiten durch Denknotwendigkeiten ersetzen.*”
>
> Albert Einstein

The two precise descriptions of motion available in the year 2000, namely that of general relativity and that of the standard model, are both useful and thoroughly beautiful. The descriptions are *useful* because their consequences are confirmed by all experiments, to the full precision that is possible at present. We are able to describe and understand all examples of motion that have ever been encountered. We can use this understanding to save lives, provide food and enjoy life. We have thus reached a considerable height in our mountain ascent. Our quest for the full description of motion is not far from completion.

The results of twentieth century physics are also *beautiful*. By this, physicists just mean that they can be phrased in *simple* terms. (This is a poor definition of beauty, but physicists are rarely experts on beauty.) In particular, both general relativity and the standard model are described by a Lagrangian, i.e., by a procedure that measures how things change. The use of a Lagrangian is possible because all motion observed in nature minimizes action. Since in physics, action is a measure of change, we can say that all motion observed in nature *minimizes change*.

On the other hand, an important aspect of any type of motion, the masses of the involved elementary particles and the strength of their coupling, are unexplained by general relativity and by the standard model of particle physics. The same applies to the origin of all the particles in the universe, their initial conditions, and the dimensionality of space-time. Obviously, the millennium description of physics is not yet complete.

This last part of our hike will be the most demanding. In the ascent of any high mountain, the head gets dizzy because of the lack of oxygen. The finite amount of energy at our disposal requires that we leave behind all unnecessary baggage and everything that slows us down. In order to determine what is unnecessary, we need to focus on what we want to achieve. Our aim is the precise description of motion. But even though general relativity and quantum theory are extremely precise, we carry a burden: the two theories and their concepts *contradict* each other. Let us focus on these contradictions.

---

* ‘One needs to replace habits of thought by necessities of thought.’

THE CONTRADICTIONS

In classical physics and in general relativity, the vacuum, or empty space, is a region with no mass, no energy and no momentum. If particles or gravitational fields are present, the energy density is not zero, space is curved and there is no vacuum. A simple way to measure the energy content of space is proposed by general relativity: we measure the average curvature of the universe. Cosmological measurements reveal an average energy density E/V of the 'vacuum' with the value of

Ref. 36
Vol. II, page 217

$$\frac{E}{V} \approx 0.5\,\mathrm{nJ/m^3}\ . \tag{69}$$

In short, cosmological data shows that the energy density of intergalactic space is not zero, but extremely small.

On the other hand, quantum field theory tells a different story on vacuum energy. A vacuum is a region with zero-point fluctuations. The energy content of a vacuum is the sum of the zero-point energies of all the fields it contains. Indeed, the Casimir effect 'proves' the reality of these zero-point energies. Their energy density is given, within one order of magnitude, by

Ref. 37
Vol. V, page 82
Vol. V, page 87

$$\frac{E}{V} = \frac{4\pi h}{c^3}\int_0^{\nu_{\max}} \nu^3 \mathrm{d}\nu = \frac{\pi\, h}{c^3}\nu_{\max}^4\ . \tag{70}$$

The approximation is valid for the case in which the cut-off frequency $\nu_{\max}$ is much larger than the rest mass m of the particles corresponding to the field under consideration. The limit considerations given above imply that the cut-off energy has to be of the order of the Planck energy $\sqrt{\hbar c^5/4G}$, about $0.6\cdot10^{19}$ GeV= 1.0 GJ. That would give a vacuum energy density of

Page 36

$$\frac{E}{V} \approx 10^{111}\,\mathrm{J/m^3}\ , \tag{71}$$

which is about 10^{120} times higher than the experimental measurement. In other words, something is slightly wrong in the calculation due to quantum field theory.

General relativity and quantum theory contradict each other in other ways. Gravity is curved space-time. Extensive research has shown that quantum field theory, which describes electrodynamics and nuclear forces, fails for situations with strongly curved space-times. In these cases the concept of 'particle' is not precisely defined. Quantum field theory cannot be extended to include gravity consistently, and thus to include general relativity. Without the concept of the particle as a discrete entity, we also lose the ability to perform perturbation calculations – and these are the only calculations possible in quantum field theory. In short, quantum theory only works because it assumes that gravity does not exist. Indeed, the gravitational constant does not appear in any consistent quantum field theory.

Ref. 38

On the other hand, general relativity neglects the commutation rules between physical quantities discovered in experiments on a microscopic scale. General relativity assumes that the classical notions of position and momentum of material objects are meaningful.

It thus ignores Planck's constant $\hbar$, and only works by neglecting quantum effects.

Vol. V, page 206 Measurements also lead to problems. In general relativity, as in classical physics, it is assumed that arbitrary precision of measurement is possible – for example, by using finer and finer ruler marks. In quantum mechanics, on the other hand, the precision of measurement is limited. The indeterminacy principle yields limits that result from the mass M of the apparatus. Vol. V, page 34

The contradictions are most evident in relation to *time*. According to relativity and classical physics, time is what is read from clocks. But quantum theory says that precise clocks do not exist, especially if gravitation is taken into account. What does 'waiting 10 minutes' mean, if the clock goes into a quantum-mechanical superposition as a result of its coupling to space-time geometry?

Similarly, general relativity implies that space and time cannot be distinguished, whereas quantum theory implies that matter does make a distinction between them. Quantum theory is a theory of – admittedly weird – local observables. In general relativity, there are no local observables, as Einstein's hole argument shows. Vol. II, page 255

The contradiction between the two theories is shown most dramatically by the failure of general relativity to describe the pair creation of particles with spin 1/2, a typical and essential quantum process. John Wheeler and others have argued that, in such a case, the topology of space necessarily has to *change*; in general relativity, however, the topology of space is fixed. Equivalently, quantum theory says that matter is made of *fermions*, which cannot be incorporated into general relativity. Ref. 39, Ref. 40, Ref. 41, Ref. 42

Ref. 43 Another striking contradiction was pointed out by Jürgen Ehlers. Quantum theory is built on point particles, and point particles move on *time-like* world lines. But following general relativity, point particles have a singularity inside their black hole horizon; and singularities always move on *space-like* world lines. The two theories thus contradict each other at smallest distances.

No description of nature that contains contradictions can lead to a unified description or to a completely correct description. In order to proceed, let us take the shortest and fastest path: let us investigate the origin of the contradictions in more detail.

THE ORIGIN OF THE CONTRADICTIONS

Ref. 44 We take the simplest way to deduce the origin of *all* contradictions between general relativity and quantum mechanics. In 20th-century physics, motion is described in terms of objects, made up of particles, and space-time, made up of events. Let us see how these two concepts are defined.

A *particle* – and in general any object – is defined as a conserved entity which has a position and which can move. In fact, the etymology of the word *object* is connected to the latter property. In other words, a particle is a small entity with conserved mass, charge, spin and so on, whose position can vary with time.

Ref. 45 In every physics text, *time* is defined with the help of moving objects, usually called 'clocks', or moving particles, such as those emitted by light sources. Similarly, *length* is defined in terms of objects, either with an old-fashioned ruler or in terms of the motion of light, which is itself motion of particles.

Modern physics has sharpened our definitions of particles and space-time. Quantum mechanics assumes that space-time is given (as a symmetry of the Hamiltonian), and

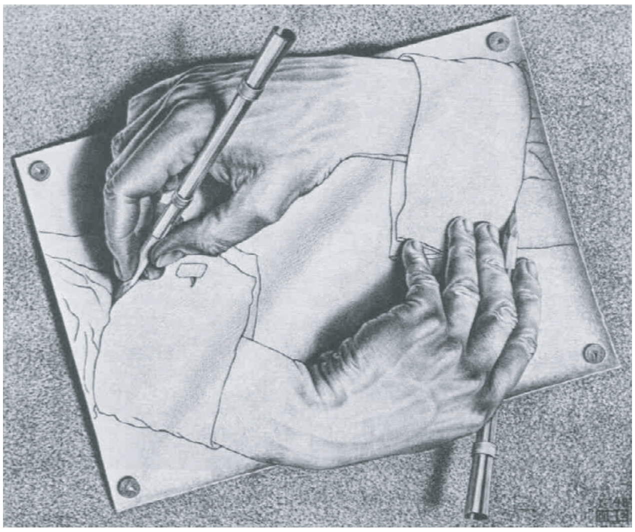


FIGURE 2 ‘Tekenen’ by Maurits Escher, 1948 – a metaphor for the way in which ‘particles’ and ‘space-time’ are usually defined: each with the help of the other (© M.C. Escher Heirs).

studies the properties of particles and their motion, both for matter and for radiation. Quantum theory has deduced the full list of properties that define a particle. General relativity, and especially cosmology, takes the opposite approach: it assumes that the properties of matter and radiation are given (for example, via their equations of state), and describes in detail the space-time that follows from them, in particular its curvature.

However, one fact remains unchanged throughout all these advances: the two concepts of particle and of space-time are *each defined with the help of the other*. To eliminate the contradiction between quantum mechanics and general relativity, and to create a more complete theory, we must eliminate this circular definition. As we will discover below, this necessitates a radical change in our description of nature, and in particular of the continuity of space-time.

For a long time, the contradictions between these two descriptions of nature were accommodated by keeping the two theories separate. One often hears the statement that quantum mechanics is valid at small scales and general relativity is valid at large scales. But this separation is artificial; worse, it prevents us from solving the problem.

The situation resembles the well-known drawing, Figure 2, by Maurits Escher (1898–1972) in which two hands, each holding a pencil, seem to be drawing each other. If one hand is taken as a symbol of vacuum and the other as a symbol of particles, with the act of drawing taken as the act of defining, the picture gives a description of twentieth-century physics. The apparent circular definition is solved by recognizing that the two concepts (the two hands) both originate from a third, hidden concept. In the picture, this third entity is the hand of the artist.

We thus conclude that the circular definition between vacuum and matter is solved by *common constituents*. In order to find out what these common constituents are and what they are not, we focus on that domain where the contradictions are most dramatic.

TABLE 2 The size, Schwarzschild radius and Compton wavelength of some objects appearing in nature. The lengths in quotation marks make no physical sense, as explained in the text.

Object	Diameter d	Mass m	Schwarzschild radius r_S	Ratio d/r_S	Compton wavelength λ_C	Ratio d/λ_C
galaxy	≈ 1 Zm	$\approx 5 \cdot 10^{40}$ kg	≈ 70 Tm	$\approx 10^{7}$	‘$\approx 10^{-83}$ m’	$\approx 10^{104}$
neutron star	10 km	$2.8 \cdot 10^{30}$ kg	4.2 km	2.4	‘$1.3 \cdot 10^{-73}$ m’	$8.0 \cdot 10^{76}$
Sun	1.4 Gm	$2.0 \cdot 10^{30}$ kg	3.0 km	$4.8 \cdot 10^{5}$	‘$1.0 \cdot 10^{-73}$ m’	$8.0 \cdot 10^{81}$
Earth	13 Mm	$6.0 \cdot 10^{24}$ kg	8.9 mm	$1.4 \cdot 10^{9}$	‘$5.8 \cdot 10^{-68}$ m’	$2.2 \cdot 10^{74}$
human	1.8 m	75 kg	0.11 ym	$1.6 \cdot 10^{25}$	‘$4.7 \cdot 10^{-45}$ m’	$3.8 \cdot 10^{44}$
molecule	10 nm	0.57 zg	‘$8.5 \cdot 10^{-52}$ m’	$1.2 \cdot 10^{43}$	$6.2 \cdot 10^{-19}$ m	$1.6 \cdot 10^{10}$
atom (^{12}C)	0.6 nm	20 yg	‘$3.0 \cdot 10^{-53}$ m’	$2.0 \cdot 10^{43}$	$1.8 \cdot 10^{-17}$ m	$3.2 \cdot 10^{7}$
proton p	2 fm	1.7 yg	‘$2.5 \cdot 10^{-54}$ m’	$8.0 \cdot 10^{38}$	$2.0 \cdot 10^{-16}$ m	9.6
pion π	2 fm	0.24 yg	‘$3.6 \cdot 10^{-55}$ m’	$5.6 \cdot 10^{39}$	$1.5 \cdot 10^{-15}$ m	1.4
up-quark u	< 0.1 fm	0.6 yg	‘$9.0 \cdot 10^{-55}$ m’	$< 1.1 \cdot 10^{38}$	$5.5 \cdot 10^{-16}$ m	< 0.18
electron e	< 4 am	$9.1 \cdot 10^{-31}$ kg	‘$1.4 \cdot 10^{-57}$ m’	$3.0 \cdot 10^{39}$	$3.8 \cdot 10^{-13}$ m	$< 1.0 \cdot 10^{-5}$
neutrino ν_e	< 4 am	$< 3.0 \cdot 10^{-35}$ kg	‘$< 4.5 \cdot 10^{-62}$ m’	n.a.	$> 1.1 \cdot 10^{-8}$ m	$< 3.4 \cdot 10^{-10}$

The place of contradiction: Planck scales

Both general relativity and quantum mechanics are successful theories for the description of nature: they agree with all data. Each theory provides a criterion for determining when they are necessary and when classical Galilean physics is no longer applicable.

General relativity shows that it is *necessary* to take into account the curvature of empty space* and space-time whenever we approach an object of mass m to within a distance of the order of the Schwarzschild radius r_S, given by

$$r_S = 2Gm/c^2 \; . \qquad (72)$$

The gravitational constant G and the speed of light c act as conversion constants. Indeed, as the Schwarzschild radius of an object is approached, the difference between general relativity and the classical $1/r^2$ description of gravity becomes larger and larger. For example, the barely measurable gravitational deflection of light by the Sun is due to the light approaching the Sun to within $2.4 \cdot 10^5$ times its Schwarzschild radius. Usually, we are forced to stay away from objects at a distance that is an even larger multiple of the Schwarzschild radius, as shown in Table 2. Only for this reason is general relativity unnecessary in everyday life. (An object smaller than its own Schwarzschild radius is called a *black hole*, because according to general relativity, no signals from inside the Schwarzschild radius can reach the outside world.)

Ref. 39, Ref. 46

Ref. 47

Similarly, quantum mechanics shows that Galilean physics must be abandoned and quantum effects *must* be taken into account whenever an object is approached to within

---

* In the following, we use the terms ‘vacuum’ and ‘empty space’ interchangeably.

distances of the order of the (reduced) Compton wavelength λ_C, given by

$$\lambda_C = \frac{\hbar}{m\,c} \;. \tag{73}$$

In this case, Planck's constant $\hbar$ and the speed of light c act as conversion factors to transform the mass m into a length scale. Of course, this length is only relevant if the object is smaller than its own Compton wavelength. At these scales we get relativistic quantum effects, such as particle–antiparticle creation or annihilation. Table 2 shows that the approach distance is near to or smaller than the Compton wavelength only in the microscopic world, so that such effects are not observed in everyday life. Only for this reason we do not need quantum field theory to describe common observations.

The *combined* concepts of quantum field theory and general relativity are required in situations where both conditions are satisfied simultaneously. The necessary approach distance for such situations is calculated by setting $r_S = 2\lambda_C$ (the factor 2 is introduced for simplicity). We find that this is the case when lengths or times are (of the order of)

$$\begin{aligned} l_{Pl} = \sqrt{\hbar G/c^3} &= 1.6 \cdot 10^{-35}\ \text{m, the } \textit{Planck length,} \\ t_{Pl} = \sqrt{\hbar G/c^5} &= 5.4 \cdot 10^{-44}\ \text{s, the } \textit{Planck time.} \end{aligned} \tag{74}$$

Whenever we approach objects at these scales, both general relativity and quantum mechanics play a role, and effects of *quantum gravity* appear. Because the values of the Planck dimensions are extremely small, this level of sophistication is unnecessary in everyday life, in astronomy and even in particle physics.

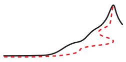
However, to answer the questions posed at the beginning – why do we live in three dimensions, why are there three interactions, and why is the proton 1836.15 times heavier than the electron? – we require a precise and complete description of nature. To answer these questions, we must understand physics at the Planck scale.

Note that the Planck scales specify one of only two domains of nature where quantum Challenge 28 s mechanics and general relativity apply simultaneously. (What is the other?) As Planck scales are the easier domain to study, they provide the best starting point. When Max Ref. 48 Planck discovered the existence of those scales, he was interested in them mainly as *natural units of measurement*. However, their importance in nature is much greater. As we will discover, they tell us about the common constituents of space and time.

Summary on the clash between the two theories

General relativity and quantum theory contradict each other. In practice, this happens only at Planck scales. The reason for the contradiction is the insistence on describing space with points and particles with point-like objects.

We must stop using points as basic elements of our description of nature. The next chapters will show that instead of points, we must introduce common constituents for space and particles.



Chapter 4

DOES MATTER DIFFER FROM VACUUM?

Vol. II, page 22

The appearance of the quantum of action in the description of motion leads to limitations for all measurements: Heisenberg's indeterminacy relations. The indeterminacy relations, when combined with the effects of gravitation, imply an almost unbelievable series of consequences. Among them is the equivalence of vacuum and matter.

Farewell to instants of time

> Time is composed of time atoms ... which in fact are indivisible.
>
> Moses Maimonides, twelfth century

Ref. 49, Ref. 50

Measurement limits appear most clearly when we investigate the properties of clocks and metre rules. Is it possible to construct a clock that is able to measure time intervals shorter than the Planck time? Surprisingly, the answer is no, even though the time–energy indeterminacy relation $\Delta E \Delta t \geqslant \hbar$ seems to indicate that by making ΔE large enough, we can make Δt arbitrary small.

Ref. 51, Ref. 52
Ref. 53

Every clock is a device with some moving parts. Parts can be mechanical wheels, particles of matter in motion, changing electrodynamic fields (i.e., photons), or decaying radioactive particles. For each moving component of a clock the indeterminacy principle applies. As explained most clearly by Michael Raymer, the indeterminacy relation for two non-commuting variables describes two different, but related, situations: it makes a statement about standard *deviations* of *separate* measurements on *many* identical systems; and it describes the measurement *precision* for a *joint* measurement on a *single* system. In what follows, we will consider only the second situation.

For a clock to be useful, we need to know both the time and the energy of each hand. Otherwise it would not be a recording device. More generally, a clock must be a classical system. We need the combined knowledge of the non-commuting variables for each moving component of the clock. Let us focus on the component with the largest time indeterminacy Δt. It is evident that the smallest time interval δt that can be measured by a clock is always larger than the quantum limit, i.e., larger than the time indeterminacy Δt for the most 'uncertain' component. Thus we have

$$\delta t \geqslant \Delta t \geqslant \frac{\hbar}{\Delta E} , \qquad (75)$$

where ΔE is the energy indeterminacy of the moving component, and ΔE must be smaller than the total energy $E = mc^2$ of the component itself.* Furthermore, a clock provides information, so signals have to be able to leave it. Therefore the clock must not be a black hole: its mass m must be smaller than the Schwarzschild mass for its size, i.e., $m \leqslant c^2 l/G$, where l is the size of the clock (neglecting factors of order unity). Finally, for a sensible measurement of the time interval δt, the size l of the clock must be smaller than $c\,\delta t$, because otherwise different parts of the clock could not work together to produce the same time display.** If we combine all these conditions, we get

$$\delta t \geqslant \frac{\hbar G}{c^5 \delta t} \tag{76}$$

or

$$\delta t \geqslant \sqrt{\frac{\hbar G}{c^5}} = t_{\mathrm{Pl}} \,. \tag{77}$$

In summary, from three simple properties of any clock – namely, that it is only a single clock, that we can read its dial, and that it gives sensible read-outs – we conclude that *clocks cannot measure time intervals shorter than the Planck time*. Note that this argument is independent of the nature of the clock mechanism. Whether the clock operates by gravitational, electrical, mechanical or even nuclear means, the limit still applies.***

Ref. 59 The same conclusion can be reached in other ways. For example, any clock small enough to measure small time intervals necessarily has a certain energy indeterminacy due to the indeterminacy relation. Meanwhile, on the basis of general relativity, any energy density induces a deformation of space-time, and signals from the deformed region arrive with a certain delay due to that deformation. (Vol. II, page 161) The energy indeterminacy of the source leads to an indeterminacy in the deformation, and thus in the delay. The expression from general relativity for the deformation of the time part of the line element due to a mass m is $\delta t = mG/lc^3$. (Ref. 46) From the mass–energy relation, we see that an energy spread ΔE produces an indeterminacy Δt in the delay:

$$\Delta t = \frac{\Delta E\, G}{l\, c^5} \,. \tag{78}$$

This determines the precision of the clock. Furthermore, the energy indeterminacy of the clock is fixed by the indeterminacy relation for time and energy $\Delta E \geqslant \hbar/\Delta t$, which is

---

* Physically, this condition means being sure that there is only *one* clock: if $\Delta E > E$, it would be impossible to distinguish between a single clock and a clock–anticlock pair created from the vacuum, or a component together with two such pairs, and so on.

** It is amusing to explore how a clock larger than $c\,\delta t$ would stop working, as a result of the loss of rigidity in its components. Challenge 29 s

*** Note that gravitation is essential here. The present argument differs from the well-known study on the limitations of clocks due to their mass and their measuring time which was published by Salecker and Wigner and summarized in pedagogical form by Zimmerman. (Ref. 54, Ref. 55) In our case, both quantum mechanics and gravity are included, and therefore a different, lower, and more fundamental limit is found. Note also that the discovery of black hole radiation does not change the argument: black hole radiation notwithstanding, measurement devices cannot exist inside black holes. (Ref. 56, Ref. 57) in turn fixed by the precision of the clock. Combining all this, we again find the relation $\delta t \geqslant t_{\mathrm{Pl}}$ for the minimum measurable time.

We are forced to conclude that *in nature there is a minimum time interval.* In other words, *at Planck scales the term 'instant of time' has no theoretical or experimental basis.*

But let us go on. Special relativity, quantum mechanics and general relativity all rely on the idea that time can be defined for all points of a given reference frame. However, two clocks a distance l apart cannot be synchronized with arbitrary precision. Since the distance between two clocks cannot be measured with an error smaller than the Planck length l_{Pl}, and transmission of signals is necessary for synchronization, it is not possible to synchronize two clocks with a better precision than $l_{\mathrm{Pl}}/c = t_{\mathrm{Pl}}$, the Planck time. So use of a single *time coordinate* for a whole reference frame is only an approximation. *Reference frames do not have a single time coordinate at Planck scales.*

Moreover, since the time difference between events can only be measured within a Planck time, for two events distant in time by this order of magnitude, it is not possible Ref. 69 to say with complete certainty which of the two precedes the other. But if events cannot be ordered, then the very concept of time, which was introduced into physics to describe sequences, makes no sense at Planck scales. In other words, after dropping the idea of a common time coordinate for a complete frame of reference, we are forced to drop the idea of time at a single 'point' as well. *The concept of 'proper time' loses its meaning at Planck scales.*

FAREWELL TO POINTS IN SPACE

> Our greatest pretenses are built up not to hide the evil and the ugly in us, but our emptyness. The hardest thing to hide is something that is not there.
>
> Eric Hoffer, *The Passionate State of Mind*

In a similar way, we can deduce that it is impossible to make a metre rule, or any other Ref. 20 length-measuring device, that is able to measure lengths shorter than the Planck length. Obviously, we can already deduce this from $l_{\mathrm{Pl}} = c\, t_{\mathrm{Pl}}$, but an independent proof is also possible.

For any length measurement, joint measurements of position and momentum are necessary. The most straightforward way to measure the distance between two points is to put an object at rest at each position. Now, the minimal length δl that can be measured must be larger than the position indeterminacy of the two objects. From the indeterminacy principle it is known that neither object's position can be determined with a precision Δl better than that given by the indeterminacy relation $\Delta l\, \Delta p = \hbar$, where Δp is the momentum indeterminacy. The requirement that there be only one object at each end (avoiding pair production from the vacuum) means that $\Delta p < mc$: together, these requirements give

$$\delta l \geqslant \Delta l \geqslant \frac{\hbar}{mc} . \tag{79}$$

Furthermore, the measurement cannot be performed if signals cannot leave the objects; thus, they cannot be black holes. Therefore their masses must be small enough for their

Schwarzschild radius $r_S = 2Gm/c^2$ to be less than the distance δl separating them. Again omitting the factor of 2, we get

$$\delta l \geqslant \sqrt{\frac{\hbar G}{c^3}} = l_{Pl} \,. \qquad (80)$$

Length measurements are limited by the Planck length.

Another way to deduce this limit reverses the roles of general relativity and quantum theory. To measure the distance between two objects, we have to localize the first object with respect to the other within a certain interval Δx. The corresponding energy indeterminacy obeys $\Delta E = c(c^2m^2 + (\Delta p)^2)^{1/2} \geqslant c\hbar/\Delta x$. However, general relativity shows that a small volume filled with energy changes the curvature of space-time, and thus changes the metric of the surrounding space. For the resulting distance change Δl, compared with empty space, we find the expression $\Delta l \approx G\Delta E/c^4$. In short, if we localize the first particle in space with a precision Δx, the distance to a second particle is known only with precision Δl. The minimum length δl that can be measured is obviously larger than either of these quantities; inserting the expression for ΔE, we find again that the minimum measurable length δl is given by the Planck length.

Ref. 39, Ref. 46

Ref. 20

Ref. 60

Ref. 61, Ref. 62

Ref. 22

We note that every length measurement requires a joint measurement of position and momentum. This is particularly obvious if we approach a metre ruler to an object, but it is equally true for any other length measurement.

We note that, since the Planck length is the shortest possible length, there can be no observations of quantum-mechanical effects for a situation where the corresponding de Broglie or Compton wavelength is smaller than the Planck length. In proton–proton collisions we observe both pair production and interference effects. In contrast, the Planck limit implies that in everyday, macroscopic situations, such as car–car collisions, we cannot observe embryo–antiembryo pair production and quantum interference effects.

Another way to convince oneself that points have no meaning is to observe that a point is an entity with vanishing volume; however, the minimum volume possible in nature is the Planck volume $V_{Pl} = l_{Pl}^3$.

We conclude that the Planck units not only provide *natural* units; they also provide – within a factor of order one – the *limit* values of space and time intervals.

In summary, from two simple properties common to all length-measuring devices, namely that they are discrete and that they can be read, we arrive at the conclusion that *lengths smaller than the Planck length cannot be measured*. Whatever method is used, be it a metre rule or time-of-flight measurement, we cannot overcome this fundamental limit. It follows that *the concept of a 'point in space' has no experimental basis*.

The limitations on length measurements imply that we cannot speak of continuous space, except in an approximate sense. As a result of the lack of measurement precision at Planck scales, the concepts of spatial order, of translation invariance, of isotropy of the vacuum and of global coordinate systems have no experimental basis.

THE GENERALIZED INDETERMINACY PRINCIPLE

Ref. 20 The limit values for length and time measurements are often expressed by the so-called *generalized indeterminacy principle*

$$\Delta p \Delta x \geqslant \hbar/2 + f \frac{G}{c^3} (\Delta p)^2 \tag{81}$$

or

$$\Delta p \Delta x \geqslant \hbar/2 + f \frac{l_{\mathrm{Pl}}^2}{\hbar} (\Delta p)^2 \;, \tag{82}$$

where f is a numerical factor of order unity. A similar expression holds for the time–energy indeterminacy relation. The first term on the right-hand side is the usual quantum-mechanical indeterminacy. The second term is negligible for everyday energies, and is significant only near Planck energies; it is due to the changes in space-time induced by gravity at these high energies. You should be able to show that the generalized principle (81) implies that Δx can never be smaller than $f^{1/2} l_{\mathrm{Pl}}$. Challenge 30 e

The generalized indeterminacy principle is derived in exactly the same way in which Heisenberg derived the original indeterminacy principle $\Delta p \Delta x \geqslant \hbar/2$, namely by studying the deflection of light by an object under a microscope. A careful re-evaluation of the process, this time including gravity, yields equation (81). For this reason, *all* descriptions that unify quantum mechanics and gravity must yield this relation, and indeed they do. Ref. 20 Ref. 63, Ref. 64 Ref. 65, Ref. 66, Ref. 67

FAREWELL TO SPACE-TIME CONTINUITY

> “One can give good reasons why reality cannot at all be represented by a continuous field. From the quantum phenomena it appears to follow with certainty that a finite system of finite energy can be completely described by a finite set of numbers (quantum numbers). This does not seem to be in accordance with a continuum theory, and must lead to an attempt to find a purely algebraic theory for the description of reality. But nobody knows how to obtain the basis of such a theory.”
>
> Albert Einstein, 1955, the last sentences of *The Meaning of Relativity – Including the Relativistic Theory of the Non-Symmetric Field*, fifth edition. These were also his last published words.

We remember that quantum mechanics begins with the realization that the classical concept of action makes no sense below the value of $\hbar/2$; similarly, unified theories begin with the realization that the classical concepts of time and length make no sense below Planck scales. However, the 20th-century description of nature does contain such small values: it involves intervals smaller than the smallest measurable one. Therefore, *the continuum description of space-time has to be abandoned* in favour of a more appropriate description.

Ref. 68 The minimum length distance, the minimum time interval, and equivalently, the new, generalized indeterminacy relation appearing at Planck scales show that space, time and in particular, space-time, are not well described as a continuum. Inserting $c\Delta p \geqslant \Delta E \geqslant \hbar/\Delta t$ into equation (81), we get

$$\Delta x \Delta t \geqslant \hbar G/c^4 = t_{\mathrm{Pl}} l_{\mathrm{Pl}} , \tag{83}$$

which of course has no counterpart in standard quantum mechanics. This shows that also space-time *events do not exist*. The concept of an 'event', being a combination of a 'point in space' and an 'instant of time', loses its meaning for the description of nature at Planck scales.

Interestingly, the view that continuity must be abandoned is almost one hundred years old. Already in 1917, Albert Einstein wrote in a letter to Werner Dällenbach:

Ref. 73

> Wenn die molekulare Auffassung der Materie die richtige (zweckmässige) ist, d.h. wenn ein Teil Welt durch eine endliche Zahl bewegter Punkte darzustellen ist, so enthält das Kontinuum der heutigen Theorie *zu viel* Mannigfaltigkeit der Möglichkeiten. Auch ich glaube, dass dieses zu viel daran schuld ist, dass unsere heutige Mittel der Beschreibung an der Quantentheorie scheitern. Die Frage scheint mir, wie man über ein Diskontinuum Aussagen formulieren kann, ohne ein Kontinuum (Raum-Zeit) zu Hilfe zu nehmen; letzteres wäre als eine im Wesen des Problems nicht gerechtfertigte zusätzliche Konstruktion, der nichts „Reales" entspricht, aus der Theorie zu verbannen. Dazu fehlt uns aber leider noch die mathematische Form. Wie viel habe ich mich in diesem Sinne schon geplagt!
>
> Allerdings sehe ich auch hier prinzipielle Schwierigkeiten. Die Elektronen (als Punkte) wären in einem solchen System letzte Gegebenheiten (Bausteine). Gibt es überhaupt letzte Bausteine? Warum sind diese alle von gleicher Grösse? Ist es befriedigend zu sagen: Gott hat sie in seiner Weisheit alle gleich gross gemacht, jedes wie jedes andere, weil er so wollte; er hätte sie auch, wenn es ihm gepasst hätte, verschieden machen können. Da ist man bei der Kontinuum-Auffassung besser dran, weil man nicht von Anfang an die Elementar-Bausteine angeben muss. Ferner die alte Frage vom Vakuum! Aber diese Bedenken müssen verblassen hinter der blendenden Tatsache: Das Kontinuum ist ausführlicher als die zu beschreibenden Dinge...
>
> Lieber Dällenbach! Was hilft alles Argumentieren, wenn man nicht bis zu einer befriedigenden Auffassung durchdringt; das aber ist verteufelt schwer. Es wird einen schweren Kampf kosten, bis man diesen Schritt, der uns da vorschwebt, wirklich gemacht haben wird. Also strengen Sie Ihr Gehirn an, vielleicht zwingen Sie es.*

---

* 'If the molecular conception of matter is the right (appropriate) one, i.e., if a part of the world is to be represented by a finite number of moving points, then the continuum of the present theory contains too great a manifold of possibilities. I also believe that this too great is responsible that our present means of description fail for quantum theory. The questions seems to me how one can formulate statements about a discontinuum without using a continuum (space-time) as an aid; the latter should be banned from the theory as a supplementary construction not justified by the essence of the problem, which corresponds to nothing "real". But unfortunately we still lack the mathematical form. How much have I already plagued myself in this direction!

The second half of this text will propose a way to rise to the challenge. At this point however, we first complete the exploration of the limitations of continuum physics.

In 20th century physics, space-time points are idealizations of events – but this idealization is inadequate. The use of the concept of 'point' is similar to the use of the concept of 'aether' a century ago: it is impossible to measure or detect, and it is only useful as a convenient notion until a way to describe nature without it has been found. *Like the 'aether', also 'points' lead reason astray.*

All paradoxes resulting from the infinite divisibility of space and time, such as Zeno's argument on the impossibility of distinguishing motion from rest, or the Banach–Tarski paradox, are now avoided. We can dismiss them straight away because of their incorrect premises concerning the nature of space and time.

The consequences of the Planck limits for measurements of time and space can be expressed in other ways. It is often said that given any two points in space or any two instants of time, there is always a third in between. Physicists sloppily call this property continuity, while mathematicians call it denseness. However, at Planck scales this property cannot hold, since there are no intervals smaller than the Planck time. Thus points and instants are not dense, and *between two points there is not always a third.* This results again means that *space and time are not continuous.* Of course, at large scales they are – approximately – continuous, in the same way that a piece of rubber or a liquid seems continuous at everyday scales, even though it is not at a small scale. But in nature, space, time and space-time are not continuous entities.

But there is more to come. The very existence of a minimum length contradicts the theory of special relativity, in which it is shown that lengths undergo Lorentz contraction when the frame of reference is changed. There is only one conclusion: special relativity (and general relativity) cannot be correct at very small distances. Thus, *space-time is not Lorentz-invariant (nor diffeomorphism-invariant) at Planck scales.* All the symmetries that are at the basis of special and general relativity are only approximately valid at Planck scales.

The imprecision of measurement implies that most familiar concepts used to describe spatial relations become useless. For example, the concept of a *metric* loses its usefulness at Planck scales, since distances cannot be measured with precision. So it is impossible to say whether space is flat or curved. The impossibility of measuring lengths exactly is equivalent to fluctuations of the curvature, and thus of gravity. Ref. 20, Ref. 70

In short, space and space-time are not smooth at Planck scales. This conclusion has important implications. For example, the conclusion implies that certain mathematical solutions found in books on general relativity, such as the Eddington–Finkelstein coor-

---

Yet I also see difficulties of principle. In such a system the electrons (as points) would be the ultimate entities (building blocks). Do ultimate building blocks really exist? Why are they all of equal size? Is it satisfactory to say: God in his wisdom made them all equally big, each like every other one, because he wanted it that way; he could also have made them, if he had wanted, all different. With the continuum viewpoint one is better off, because one doesn't have to prescribe elementary building blocks from the outset. Furthermore, the old question of the vacuum! But these considerations must pale beside the dazzling fact: The continuum is more ample than the things to be described...

Dear Dällenbach! All arguing does not help if one does not achieve a satisfying conception; but this is devilishly difficult. It will cost a difficult fight until the step that we are thinking of will be realized. Thus, squeeze your brain, maybe you can force it.'

dinates and the Kruskal–Szekeres coordinates do *not* describe nature! Indeed, these coordinate systems, which claim to show that space-time goes on *behind* the horizon of a black hole, are based on the idea that space-time is smooth everywhere. However, quantum physics shows that space-time is not smooth at the horizon, but fluctuates wildly there. In short, quantum physics confirms what common sense already knew: behind a horizon, nothing can be observed, and thus there is nothing there.

Farewell to dimensionality

Even the number of spatial dimensions makes no sense at Planck scales. Let us remind ourselves how to determine this number experimentally. One possible way is to determine how many points we can choose in space such that all the distances between them are equal. If we can find at most n such points, the space has $n-1$ dimensions. But if reliable length measurement at Planck scales is not possible, there is no way to determine reliably the number of dimensions of space with this method.

Another way to check for three spatial dimensions is to make a knot in a shoe string and glue the ends together: since it stays knotted, we know that space has three dimensions, because there is a mathematical theorem that in spaces with greater or fewer than three dimensions, knots do not exist. Again, at Planck scales, we cannot say whether a string is knotted or not, because measurement limits at crossings make it impossible to say which strand lies above the other.

There are many other methods for determining the dimensionality of space.* In all cases, the definition of dimensionality is based on a precise definition of the concept of neighbourhood. At Planck scales, however, length measurements do not allow us to say whether a given point is inside or outside a given region. In short, whatever method we use, the lack of precise length measurements means that *at Planck scales, the dimensionality of physical space is not defined.*

Farewell to the space-time manifold

> “There is nothing in the world but matter in motion, and matter in motion cannot move otherwise than in space and time.”
>
> Lenin

The reasons for the problems with space-time become most evident when we remember Ref. 71 Euclid's well-known definition: 'A point is that which has no part.' As Euclid clearly understood, a *physical* point, as an idealization of position, cannot be defined without some measurement method. *Mathematical* points, however, can be defined without reference to a metric. They are just elements of a set, usually called a 'space'. (A 'measurable' or 'metric' space is a set of points equipped with a measure or a metric.)

---

* For example, we can determine the dimension using only the topological properties of space. If we draw a so-called *covering* of a topological space with open sets, there are always points that are elements of several sets of the covering. Let p be the maximal number of sets of which a point can be an element in a given covering. The minimum value of p over all possible coverings, minus one, gives the dimension of the space.

In fact, if physical space is not a manifold, the various methods for determining the dimensionality may give different answers. Indeed, for linear spaces without norm, the dimensionality cannot be defined in a unique way. Different definitions (fractal dimension, Lyapunov dimension, etc.) are possible.

In the case of physical space-time, the concepts of measure and of metric are more fundamental than that of a point. Confusion between physical and mathematical space and points arises from the failure to distinguish a mathematical metric from a physical length measurement.*

Vol. I, page 53 Ref. 72 Perhaps the most beautiful way to make this point is the Banach–Tarski theorem, which clearly shows the limits of the concept of volume. The theorem states that a sphere made up of *mathematical points* can be cut into five pieces in such a way that the pieces can be put together to form two spheres, each of the same volume as the original one. However, the necessary 'cuts' are infinitely curved and detailed: the pieces are wildly disconnected. For physical matter such as gold, unfortunately – or fortunately – the existence of a minimum length, namely the atomic distance, makes it impossible to perform such a cut. For vacuum, the puzzle reappears. For example, the energy of zero-point fluctuations is given by the density times the volume; following the Banach–Tarski theorem, the zero-point energy content of a single sphere should be equal to the zero-point energy of two similar spheres each of the same volume as the original one. The paradox is resolved by the Planck length, which provides a fundamental length scale even for vacuum, thus making infinitely complex cuts impossible. Therefore, the concept of volume is only well defined at Planck scales if a minimum length is introduced.

To sum up, *physical space-time cannot be a set of mathematical points.*

But there are more surprises. At Planck scales, since both temporal and spatial order break down, there is no way to say if the distance between two nearby space-time regions is space-like or time-like. *At Planck scales, time and space cannot be distinguished from each other*.

In addition, we cannot state that the topology of space-time is fixed, as general relativity implies. The topology changes mentioned above that are required for particle reactions do become possible. In this way another of the contradictions between general relativity and quantum theory is resolved.

In summary, space-time at Planck scales is not continuous, not ordered, not endowed with a metric, not four-dimensional, and not made up of points. It satisfies none of the defining properties of a manifold.** We conclude that *the concept of a space-time manifold has no justification at Planck scales*. This is a strong result. Even though both general relativity and quantum mechanics use continuous space-time, the combined theory does not.

---

* Where does the incorrect idea of continuous space-time have its roots? In everyday life, as well as in physics, space-time is a book-keeping device introduced to describe observations. Its properties are extracted from the properties of observables. Since observables can be added and multiplied, like numbers, we infer that they can take continuous values, and, in particular, arbitrarily small values. It is then possible to define points and sets of points. A special field of mathematics, topology, shows how to start from a set of points and construct, with the help of neighbourhood relations and separation properties, first a *topological space*, then, with the help of a metric, a *metric space*. With the appropriate compactness and connectedness relations, a *manifold*, characterized by its dimension, metric and topology, can be constructed.

Vol. V, page 277 ** A manifold is what looks *locally* like a Euclidean space. The exact definition can be found in the previous volume.

Farewell to observables and measurements

If space and time are not continuous, no quantities defined as derivatives with respect to space or time are precisely defined. Velocity, acceleration, momentum, energy and so on are only well defined under the assumption of continuity. That important tool, the evolution equation, is based on derivatives and can thus no longer be used. Therefore the Schrödinger and Dirac equations lose their basis. Concepts such as 'derivative', 'divergence-free' and 'source free' lose their meaning at Planck scales.

All physical observables are defined using length and time measurements. Each physical unit is a product of powers of length and time (and mass) units. (In the SI system, electrical quantities have a separate base quantity, the ampere, but the argument still holds: the ampere is itself defined in terms of a force, which is measured using the three base units of length, time and mass.) Since time and length are not continuous, *at Planck scales, observables cannot be described by real numbers.*

In addition, if time and space are not continuous, the usual expression for an observable field, $A(t, x)$, does not make sense: we have to find a more appropriate description. *Physical fields cannot exist at Planck scales.*

The consequences for quantum mechanics are severe. It makes no sense to define multiplication of observables by real numbers, but only by a discrete set of numbers. Among other implications, this means that observables do not form a linear algebra. *Observables are not described by operators at Planck scales.*

But quantum mechanics is based on the superposition principle. Without it, everything comes crumbling down. In particular, the most important observables are the gauge potentials. Since they do not form an algebra, *gauge symmetry is not valid at Planck scales*. Even innocuous-looking expressions such as $[x_i, x_j] = 0$ for $x_i \neq x_j$, which are at the root of quantum field theory, become meaningless at Planck scales. Since at those scales the superposition principle cannot be backed up by experiment, even the famous Wheeler–DeWitt equation, often assumed to describe quantum gravity, cannot be valid.

Similarly, permutation symmetry is based on the premise that we can distinguish two points by their coordinates, and then exchange particles between those locations. As we have just seen, this is not possible if the distance between the two particles is very small. We conclude that *permutation symmetry has no experimental basis at Planck scales*.

Even discrete symmetries, like charge conjugation, space inversion and time reversal, cannot be correct in this domain, because there is no way to verify them exactly by measurement. *CPT symmetry is not valid at Planck scales.*

Finally we note that all types of scaling relations break down at small scales. As a result, *renormalization symmetry is also lost at Planck scales.*

All these results are consistent: if there are no symmetries at Planck scales, there are also no observables, since physical observables are representations of symmetry groups. In fact, the limitations on time and length measurements imply that *the concept of measurement has no significance at Planck scales.*

Can space-time be a lattice?

Let us take a breath. Can a space-time lattice be an alternative to continuity?

Ref. 74 Discrete models of space-time have been studied since the 1940s. Recently, the idea Ref. 75 that space-time could be described as a lattice – like a crystal – has been explored most

Ref. 76 notably by David Finkelstein and by Gerard 't Hooft. The idea of space-time as a lattice is based on the idea that, if there is a minimum distance, then all distances are multiples Ref. 77 of this minimum.

In order to get an isotropic and homogeneous situation for large, everyday scales, the Ref. 78 structure of space-time cannot be periodic, but must be *random*. But not only must it be random in space, it must also be *fluctuating in time*. In fact, any fixed structure for space-time would violate the result that there are no lengths smaller than the Planck length: as a result of the Lorentz contraction, any moving observer would find lattice distances smaller than the Planck value. Worse still, the fixed lattice idea conflicts with general relativity, in particular with the diffeomorphism-invariance of the vacuum.

Thus, *space-time cannot be a lattice*. A minimum distance does exist in nature; however, we cannot hope that all other distances are simple multiples of it. We will discover more evidence for this negative conclusion later on.

But in fact, many discrete models of space and time have a much bigger limitation. Any such model has to answer a simple question: Where is a particle be *during* the jump from one lattice point to the next? This simple question eliminates most naive space-time models.

A GLIMPSE OF QUANTUM GEOMETRY

Given that space-time is not a set of points or events, it must be something else. We have three hints at this stage. The first is that in order to improve our description of motion we must abandon 'points', and with them, abandon the *local* description of nature. Both quantum mechanics and general relativity assume that the phrase 'observable at a point' has a precise meaning. Because it is impossible to describe space as a manifold, this expression is no longer useful. The unification of general relativity and quantum physics forces the adoption of a *non-local* description of nature at Planck scales. This is the first hint.

The existence of a minimum length implies that there is no way to physically distinguish between locations that are even closer together. We are tempted to conclude that *no* pair of locations can be distinguished, even if they are one metre apart, since on any path joining two points, no two locations that are close together can be distinguished. The problem is similar to the question about the size of a cloud or of an atom. If we measure water density or electron density, we find non-vanishing values at any distance from the centre of the cloud or the atom; however, an effective size can still be defined, because it is very unlikely that the effects of the presence of a cloud or of an atom can be seen at distances much larger than this effective size. Similarly, we can guess that two points in space-time at a macroscopic distance from each other can be distinguished because the probability that they will be confused drops rapidly with increasing distance. In short, we are thus led to a *probabilistic* description of space-time. This is the second hint. Space-time becomes a macroscopic observable, a *statistical* or *thermodynamic limit* of some microscopic entities. This is our second hint.

We note that a fluctuating structure for space-time also avoids the problems of fixed structures with Lorentz invariance. In summary, the experimental observations of special relativity – Lorentz invariance, isotropy and homogeneity – together with the notion of a minimum distance, point towards a description of space-time as *fluctuating*. This is the third hint.

Ref. 23 Several independent research efforts in quantum gravity have independently confirmed that a *non-local* and *fluctuating* description of space-time at Planck scales resolves the contradictions between general relativity and quantum theory. These are our first results on quantum geometry. To clarify the issue, we turn to the concept of the particle.

Farewell to point particles

In every example of motion, some object is involved. One of the important discoveries of the natural sciences was that all objects are composed of small constituents, called *elementary particles*. Quantum theory shows that all composite, non-elementary objects have a finite, non-vanishing size. This property allows us to determine whether a particle is elementary or not. If it behaves like a point particle, it is elementary. At present, only the leptons (electron, muon, tau and the neutrinos), the quarks and the radiation quanta of the electromagnetic, weak and strong nuclear interactions (the photon, the W and Z bosons, and the gluons) have been found to be elementary. Vol. V, page 191 A few more elementary particles are predicted by various refinements of the standard model. Protons, atoms, molecules, cheese, people, galaxies and so on are all composite, as shown in Table 2. Page 56 Elementary particles have vanishing size and are characterized by their spin and their mass.

Although the definition of 'elementary particle' as point particle is all we need in the following argument, it is not complete. It seems to leave open the possibility that future experiments could show that electrons or quarks are not elementary. This is not so! Vol. IV, page 90 In fact, any particle smaller than its own Compton wavelength is elementary. If it were composite, there would be a lighter particle inside it, which would have a larger Compton wavelength than the composite particle. This is impossible, since the size of a composite particle must be larger than the Compton wavelength of its components. (The alternative possibility that all components are heavier than the composite does not lead to satisfying physical properties: for example, it leads to intrinsically unstable components.) Ref. 79

The *size* of an object, such as those given in Table 2, is defined as the length at which differences from point-like behaviour are observed. The size d of an object is determined by measuring how it scatters a beam of probe particles. For example, the radius of the atomic nucleus was determined for the first time in Rutherford's experiment using alpha particle scattering. In daily life as well, when we look at objects, we make use of scattered photons. In general, in order for scattering to be useful, the effective wavelength $\lambda = \hbar/mv$ of the probe must be smaller than the object size d to be determined. We thus need $d > \lambda = \hbar/mv \geqslant \hbar/mc$. In addition, in order for a scattering experiment to be possible, the object must not be a black hole, since, if it were, it would simply swallow the approaching particle. This means that its mass m must be smaller than that of a black hole of the same size; in other words, from equation (72) we must have $m < dc^2/G$. Combining this with the previous condition we get

$$d > \sqrt{\frac{\hbar G}{c^3}} = l_{\mathrm{Pl}} \,. \qquad (84)$$

In other words, there is no way to observe that an object is smaller than the Planck length.

FIGURE 3 Andrei Sakharov (1921–1989).

Thus, in principle *there is thus no way to deduce from observations that a particle is point-like*. The term 'point particle' makes no sense at all! Of course, there is a relation between the existence of a minimum length for empty space and the existence of a minimum length for objects. If the term 'point of space' is meaningless, then the term 'point particle' is also meaningless. And again, the lower limit on particle size results from the combination of quantum theory and general relativity.*

Another property connected with the size of a particle is its electric dipole moment. This describes the deviation of its charge distribution from spherical. Some predictions from the standard model of elementary particles give as an *upper* limit for the electron dipole moment d_e a value of Ref. 81

$$\frac{|d_e|}{e} < 10^{-39}\,\mathrm{m}\,, \tag{85}$$

where e is the charge of the electron. This value is ten thousand times smaller than the Planck length l_{Pl}. Since the Planck length is the smallest possible length, we seem to have a contradiction here. However, a more careful, recent prediction from the standard model only states Ref. 82

$$\frac{|d_e|}{e} < 3\cdot 10^{-21}\,\mathrm{m}\,, \tag{86}$$

which is not in contradiction with our minimal length. The issue is still not settled. We will see below that these predictions are expected to be experimentally testable in the foreseeable future.

Farewell to particle properties

Planck scales have other strange consequences. In quantum field theory, the difference between a virtual particle and a real particle is that a real particle is 'on shell', obeying $E^2 = m^2c^4 + p^2c^2$, whereas a virtual particle is 'off shell'. Because of the fundamental limits of measurement precision, *at Planck scales we cannot determine whether a particle is real or virtual.*

That is not all. Antimatter can be described as matter moving backwards in time.

---

* We note that the existence of a minimum size for a particle has nothing to do with the impossibility, in quantum theory, of localizing a particle to within less than its Compton wavelength.

Since the difference between backwards and forwards cannot be determined at Planck scales, *matter and antimatter cannot be distinguished at Planck scales.*

Every particle is characterized by its spin. Spin describes two properties of a particle: its behaviour under rotations (and thus, if the particle is charged, its behaviour in magnetic fields) and its behaviour under particle exchange. The wave function of a particle with spin 1 remains invariant under a rotation of 2π, whereas that of a particle with spin 1/2 changes sign. Similarly, the combined wave function of two particles with spin 1 does not change sign under exchange of particles, whereas for two particles with spin 1/2 it does.

We see directly that both transformations are impossible to study at Planck scales. Given the limit on position measurements, the position of a rotation axis cannot be well defined, and rotations become impossible to distinguish from translations. Similarly, positional imprecision makes it impossible to determine precise separate positions for exchange experiments; at Planck scales it is impossible to say whether particle exchange has taken place or not, and whether the wave function has changed sign or not. In short, *at Planck scales, spin cannot be defined or measured, and neither fermion nor boson behaviour can be defined or measured.* In particular, this implies that supersymmetry cannot be valid at Planck scales.

Due to measurement limitations, also spatial parity cannot be defined or measured at Planck scales. Challenge 31 e

We have thus shown that at Planck scales, particles do not interact locally, are not point-like, cannot be distinguished from antiparticles, cannot distinguished from virtual particles, have no definite spin and have no definite spatial parity. We conjecture that *particles do not exist at Planck scales.* Let us explore the remaining concept: particle mass.

A mass limit for elementary particles

The size d of any elementary particle must by definition be smaller than its own (reduced) Compton wavelength $\hbar/mc$. Moreover, the size of a particle is always larger than the Planck length: $d > l_{\mathrm{Pl}}$. Combining these two requirements and eliminating the size d, we get a constraint on the mass m of any elementary particle, namely

$$m < \frac{\hbar}{c\, l_{\mathrm{Pl}}} = \sqrt{\frac{\hbar c}{G}} = m_{\mathrm{Pl}} = 2.2 \cdot 10^{-8}\,\mathrm{kg} = 1.2 \cdot 10^{19}\,\mathrm{GeV}/c^2 \,. \qquad (87)$$

The limit m_{Pl}, the so-called *Planck mass*, corresponds roughly to the mass of a human embryo that is ten days old, or equivalently, to that of a small flea. In short, *the mass of any elementary particle must be smaller than the Planck mass.* This fact was already noted as 'well known' by Andrei Sakharov* in 1968; he explains that these hypothetical particles are sometimes called 'maximons'. And indeed, the known elementary particles Ref. 24

---

* Andrei Dmitrievich Sakharov, Soviet nuclear physicist (1921–1989). One of the keenest thinkers in physics, Sakharov, among others, invented the Tokamak, directed the construction of nuclear bombs, and explained the matter–antimatter asymmetry of nature. Like many others, he later campaigned against nuclear weapons, a cause for which he was put into jail and exile, together with his wife, Yelena Bonner. He received the Nobel Peace Prize in 1975.

all have masses well below the Planck mass. (In fact, the question why their masses are so very much smaller than the Planck mass is one of the most important questions of high-energy physics. We will come back to it.)

There are many other ways to arrive at the mass limit for particles. For example, in order to measure mass by scattering – and that is the only way for very small objects – the Compton wavelength of the scatterer must be larger than the Schwarzschild radius; otherwise the probe will be swallowed. Inserting the definitions of the two quantities and neglecting the factor 2, we again get the limit $m < m_{\mathrm{Pl}}$. In fact it is a general property of descriptions of nature that a minimum space-time interval leads to an upper limit for masses of elementary particles. Ref. 80

Farewell to massive particles – and to massless vacuum

The Planck mass divided by the Planck volume, i.e., the Planck density, is given by

$$\rho_{\mathrm{Pl}} = \frac{c^5}{G^2 \hbar} = 5.2 \cdot 10^{96}\ \mathrm{kg/m}^3 \qquad (88)$$

and is a useful concept in the following. One way to measure the (gravitational) mass M enclosed in a sphere of size R, and thus (roughly) of volume R^3, is to put a test particle in orbit around it at that same distance R. Universal gravitation then gives for the mass M the expression $M = Rv^2/G$, where v is the speed of the orbiting test particle. From $v < c$, we deduce that $M < c^2 R/G$; since the minimum value for R is the Planck distance, we get (again neglecting factors of order unity) a limit for the mass density ρ, namely Challenge 32 e

$$\rho < \rho_{\mathrm{Pl}} \,. \qquad (89)$$

In other words, *the Planck density is the maximum possible value for mass density.*

Interesting things happen when we try to determine the error ΔM of a mass measurement in a Planck volume. Let us return to the mass measurement by an orbiting probe. From the relation $GM = rv^2$ we deduce by differentiation that $G\Delta M = v^2 \Delta r + 2vr\Delta v > 2vr\Delta v = 2GM\Delta v/v$. For the error Δv in the velocity measurement we have the indeterminacy relation $\Delta v \geqslant \hbar/m\Delta r + \hbar/MR \geqslant \hbar/MR$. Inserting this in the previous inequality, and again forgetting the factor of 2, we find that the mass measurement error ΔM of a mass M enclosed in a volume of size R is subject to the condition

$$\Delta M \geqslant \frac{\hbar}{cR} \,. \qquad (90)$$

Note that for everyday situations, this error is extremely small, and other errors, such as the technical limits of the balance, are much larger.

To check this result, we can explore another situation. We even use relativistic expressions, in order to show that the result does not depend on the details of the situation or the approximations. Imagine having a mass M in a box of size R, and weighing the box with a scale. (It is assumed that either the box is massless or that its mass is subtracted by the scale.) The mass error is given by $\Delta M = \Delta E/c^2$, where ΔE is due to the indeterminacy

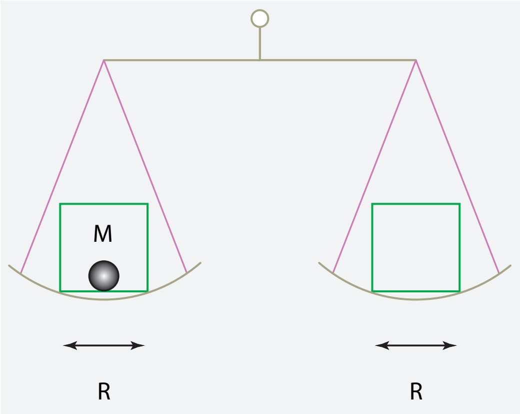


FIGURE 4 A *Gedanken experiment* showing that at Planck scales, matter and vacuum cannot be distinguished.

in the kinetic energy of the mass inside the box. Using the expression $E^2 = m^2c^4 + p^2c^2$, we get that $\Delta M \geqslant \Delta p/c$, which again reduces to equation (90). Now that we are sure of the result, let us continue.

From equation (90) we deduce that for a box of Planck dimensions, the mass measurement *error* is given by the Planck mass. But from above we also know that the mass that can be put inside such a box must not be larger than the Planck mass. Therefore, for a box of Planck dimensions, the mass measurement *error* is larger than (or at best equal to) the mass contained in it: $\Delta M \geqslant M_{\text{Pl}}$. In other words, if we build a balance with two boxes of Planck size, one empty and the other full, as shown in Figure 4, nature cannot decide which way the balance should hang! Note that even a repeated or a continuous measurement will not resolve the situation: the balance will change inclination at random, staying horizontal on average.

The argument can be rephrased as follows. The largest mass that we can put in a box of size R is a black hole with a Schwarzschild radius of the same value; the smallest mass present in such a box – corresponding to what we call a vacuum – is due to the indeterminacy relation and is given by the mass with a Compton wavelength that matches the size of the box. In other words, inside any box of size R we have a mass m, the limits of which are given by:

$$\underset{\text{(full box)}}{\frac{c^2R}{G}} \geqslant m \geqslant \underset{\text{(empty box)}}{\frac{\hbar}{cR}} . \qquad (91)$$

We see directly that for sizes R of the order of the Planck scale, the two limits coincide; in other words, we cannot distinguish a full box from an empty box in that case.

To be sure of this strange result, we check whether it also occurs if, instead of measuring the gravitational mass, as we have just done, we measure the inertial mass. The inertial mass for a small object is determined by touching it: physically speaking, by performing a scattering experiment. To determine the inertial mass inside a region of size R, a probe must have a wavelength smaller than R, and a correspondingly high energy. A high energy means that the probe also attracts the particle through gravity. (We thus find the intermediate result that *at Planck scales, inertial and gravitational mass cannot be distinguished.* Even the balance experiment shown in Figure 4 illustrates this: at Planck scales, the two types of mass are always inextricably linked.) Now, in any scattering experiment, for example in a Compton-type experiment, the mass measurement is performed by measuring the wavelength change $\delta\lambda$ of the probe before and after the scattering. The mass indeterminacy is given by

$$\frac{\Delta M}{M} = \frac{\Delta\delta\lambda}{\delta\lambda} \, . \qquad (92)$$

In order to determine the mass in a Planck volume, the probe has to have a wavelength of the Planck length. But we know from above that there is always a minimum wavelength indeterminacy, given by the Planck length l_{Pl}. In other words, for a Planck volume the wavelength error – and thus the mass error – is always as large as the Planck mass itself: $\Delta M \geqslant M_{\text{Pl}}$. Again, this limit is a direct consequence of the limit on length and space measurements.

This result has an astonishing consequence. In these examples, the measurement error is independent of the mass of the scatterer: it is the same whether or not we start with a situation in which there is a particle in the original volume. We thus find that in a volume of Planck size, it is impossible to say whether or not there is something there when we probe it with a beam!

MATTER AND VACUUM ARE INDISTINGUISHABLE

We can put these results in another way. On the one hand, if we measure the mass of a piece of vacuum of size R, the result is always at least $\hbar/cR$: there is no possible way to find a perfect vacuum in an experiment. On the other hand, if we measure the mass of a particle, we find that the result is size-dependent: at Planck scales it approaches the Planck mass for every type of particle, be it matter or radiation.

To use another image, when two particles approach each other to a separation of the order of the Planck length, the indeterminacy in the length measurements makes it impossible to say whether there is something or nothing between the two objects. In short, *matter and vacuum are interchangeable at Planck scales.* This is an important result: since mass and empty space cannot be differentiated, we have confirmed that they are made of the same 'fabric'. This idea, already suggested above, is now common to all attempts to find a unified description of nature.

This approach is corroborated by attempts to apply quantum mechanics in highly curved space-time, where a clear distinction between vacuum and particles is impossible, as shown by Fulling–Davies–Unruh radiation. Any accelerated observer, and any observer in a gravitational field, detects particles hitting him, even if he is in a vacuum. The effect shows that for curved space-time the idea of vacuum as particle-free space does not work. Since at Planck scales it is impossible to say whether or not space is flat, it is impossible to say whether it contains particles or not.

Ref. 83

Vol. V, page 87

In short, all arguments lead to the same conclusion: *vacuum, i.e., empty space-time, cannot be distinguished from matter at Planck scales.* Another common way to express this state of affairs is to say that when a particle of Planck energy travels through space it will be scattered by the fluctuations of space-time itself, as well as by matter, and the two

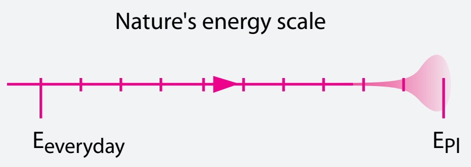


FIGURE 5 Planck effects make the energy axis an approximation.

cases are indistinguishable. These surprising results rely on a simple fact: whatever definition of mass we use, it is always measured via combined length and time measurements. (This is even the case for normal weighing scales: mass is measured by the displacement of some part of the machine.) *Mass measurement is impossible at Planck scales.* The error in such mass measurements makes it *impossible to distinguish vacuum from matter.*

Curiosities and fun challenges on Planck scales

Observers are made of matter. Observers are not made of radiation. Observers are not made of vacuum. Observers are thus biased, because they take a specific standpoint. But at Planck scale, vacuum, radiation and matter cannot be distinguished. Two results follow: first, only at those scales would a description be free of any bias in favour of matter; but secondly, observers do not exist at all at Planck energy.

* *

If measurements become impossible near Planck energy, we cannot even draw a diagram with an energy axis reaching that value. A way out is shown Figure 5. The energy of elementary particles cannot reach the Planck energy.

* *

By the standards of particle physics, the Planck energy is rather large. Suppose we wanted to impart this amount of energy to protons using a particle accelerator. How large would a *Planck accelerator* have to be? Challenge 33 s

* *

By the standards of everyday life, the Planck energy is rather small. Measured in litres of gasoline, how much fuel does it correspond to? Challenge 34 s

* *

The usual concepts of matter and of radiation are not applicable at Planck scales. Usually, it is assumed that matter and radiation are made up of interacting elementary particles. The concept of an elementary particle implies an entity that is discrete, point-like, real and not virtual, has a definite mass and a definite spin, is distinct from its antiparticle, and, most of all, is distinct from vacuum, which is assumed to have zero mass. All these properties are lost at Planck scales. *At Planck scales, the concepts of 'mass', 'vacuum', 'elementary particle', 'radiation' and 'matter' do not make sense.*

* *

Do the large errors in mass measurements imply that mass can be negative at Planck

Challenge 35 s energy?

* *

We now have a new answer to the old question: why is there something rather than nothing? At Planck scales, there is *no difference* between something and nothing. We can now honestly say about ourselves that we are made of nothing.

* *

Special relativity implies that no length or energy can be invariant. Since we have come to the conclusion that the Planck energy and the Planck length are invariant, it appears that there must be deviations from Lorentz invariance at high energy. What effects would follow? What kind of experiment could measure them? If you have a suggestion, publish it! Several attempts are being explored. We will settle the issue later on, with some interesting insights. Challenge 36 r Ref. 84 Page 248

* *

Quantum mechanics alone gives, via the Heisenberg indeterminacy relation, a lower limit to the *spread* of measurements, but, strangely enough, not on their *precision*, i.e., not on the number of significant digits. Wolfgang Jauch gives an example: atomic lattice constants are known to a much higher precision than the positional indeterminacy of single atoms inside the crystal. Ref. 85 Ref. 66

It is sometimes claimed that measurement indeterminacies smaller than the Planck values are possible for large enough numbers of particles. Can you show why this is incorrect, at least for space and time? Challenge 37 ny

* *

Of course, the idea that vacuum is not empty is not new. More than two thousand years ago, Aristotle argued for a filled vacuum, although his arguments were incorrect as seen from today's perspective. Also in the fourteenth century there was much discussion on whether empty space was composed of indivisible entities, but the debate died down again.

* *

A Planck-energy particle falling in a gravitational field would gain energy. But the Planck energy is the highest energy in nature. What does this apparent contradiction imply? Challenge 38 s

* *

One way to generalize the results presented here is to assume that, at Planck energy, nature is *event-symmetric*, i.e., symmetric under exchange of any two events. This idea, developed by Phil Gibbs, provides an additional formulation of the strange behaviour of nature at extreme scales. Ref. 59

* *

Vol. II, page 247 Because there is a minimum length in nature, so-called *singularities* do not exist. The issue, hotly debated for decades in the twentieth century, thus becomes uninteresting.

* *

Vol. V, page 102 Ref. 86 Since mass and energy density are limited, any object of finite volume has only a finite number of degrees of freedom. Calculation of the entropy of black holes has already shown us that entropy values are always finite. This implies that perfect *baths* do not exist. Baths play an important role in thermodynamics (which must therefore be viewed as only an approximation), and also in recording and measuring devices: when a device measures, it switches from a neutral state to a state in which it shows the result of the measurement. In order not to return to the neutral state, the device must be coupled to a bath. Without a bath, a reliable measuring device cannot exist. In short, perfect clocks and length-measuring devices do not exist, because nature puts a limit on their storage ability.

* *

Vol. I, page 25 If vacuum and matter cannot be distinguished, we cannot distinguish between objects and their environment. However, this was one the starting points of our journey. Some interesting adventures still await us!

* *

Vol. III, page 231 We have seen earlier that characterizing nature as made up of particles and vacuum creates problems when interactions are included. On the one hand interactions are the difference between the parts and the whole, while on the other hand, according to quantum theory, interactions are exchanges of particles. This apparent contradiction can be used to show that something is counted twice in the usual characterization of nature. However, when matter and space-time are both made of the same constituents the contradiction is resolved.

* *

Challenge 39 d Is there a smallest possible momentum? And a smallest momentum error?

* *

Challenge 40 s There is a maximum acceleration in nature. Can you deduce the value of this so-called Planck acceleration? Does it require quantum theory?

* *

Given that time becomes an approximation at Planck scales, can we still ask whether nature is *deterministic*?

Let us go back to the basics. We can define time, because in nature change is not random, but gradual. What is the situation now that we know that time is only approximate? Is non-gradual change possible? Is energy conserved? In other words, are surprises possible in nature?

It is correct to say that time is not defined at Planck scales, and that therefore that determinism is an undefinable concept, but it is not a satisfying answer. What happens at 'everyday' scales? One answer is that at our everyday scales, the probability of surprises is so small that the world indeed is effectively deterministic. In other words, nature is not really deterministic, but the departure from determinism is not measurable, since every measurement and observation, by definition, *implies* a deterministic world. The lack of surprises would be due to the limitations of our human nature – more precisely, of our senses and brain.

Challenge 41 s Can you imagine any other possibility? In truth, it is not possible to prove these answers at this point, even though the rest of our adventure will do so. We need to keep any possible alternative in mind, so that we remain able to check the answers. Page 338

* *

If matter and vacuum cannot be distinguished, then each has the properties of the other. For example, since space-time is an extended entity, matter and radiation are also extended entities. Furthermore, as space-time is an entity that reaches the borders of the system under scrutiny, particles must also do so. This is our first hint at the *extension of matter*; we will examine this argument in more detail shortly. Page 119

* *

The impossibility of distinguish matter and vacuum implies a lack of information at planck scales. In turn, this implies an intrinsic basic entropy associated with any part of the universe at Planck scales. We will come back to this topic shortly, when we discuss the entropy of black holes. Page 255

* *

Challenge 42 s When *can* matter and vacuum be distinguished? At what energy? This issue might be compared to the following question: Can we distinguish between a liquid and a gas by looking at a single atom? No, only by looking at many. Similarly, we cannot distinguish between matter and vacuum by looking at one point, but only by looking at many. We must always *average*. However, even averaging is not completely successful. Distinguishing matter from vacuum is like distinguishing clouds from the clear sky: like clouds, matter has no precise boundary.

* *

We have found that space and time are neither continuous nor made up of points. Indeed, the combination of relativity and quantum theory makes this impossible. In order to proceed in our ascent of Motion Mountain, we need to leave behind the usual concept of space-time. *At Planck scales, the concepts of 'space-time point' and 'mass point' are not applicable.*

* *

If the dimensionality of space is undefined at Planck scale, what does this mean for superstrings? Challenge 43 e

* *

Since vacuum, particles and fields are indistinguishable at Planck scales, at those scales we also lose the distinction between states and permanent, intrinsic properties of physical systems. This is a strong statement: the distinction was the starting point of our exploration of motion; the distinction allowed us to distinguish systems from their environment. Vol. I, page 27

In other words, at Planck scales we cannot talk about motion. This is a strong statement. But it is not unexpected. We are searching for the origin of motion, and we are prepared to encounter such difficulties.

Common constituents

> Ich betrachte es als durchaus möglich, dass die Physik nicht auf dem Feldbegriff begründet werden kann, d.h. auf kontinuierlichen Gebilden. Dann bleibt von meinem ganzen Luftschloss inklusive Gravitationstheorie nichts bestehen.*
>
> Albert Einstein, 1954 in a letter to Michele Besso.

> Es ist allerdings darauf hingewiesen worden, dass bereits die Einführung eines raum-zeitlichen Kontinuums angesischts der molekularen Struktur allen Geschehens im Kleinen möglicherweise als naturwidrig anzusehen sei. Vielleicht weise der Erfolg von Heisenbergs Methode auf eine rein algebraische Methode der Naturbeschreibung, auf die Ausschaltung kontinuierlicher Funktionen aus der Physik hin. Dann aber muss auch auf die Verwendung des Raum-Zeit-Knotinuums prinzipiell verzichtet werden. Es ist nicht undenkbar, dass der menschliche Scharfsinn einst Methoden finden wird, welche die Beschreitung dieses Weges möglich machen. Einstweilen aber erscheint dieses Projekt ähnlich dem Versuch, in einem luftleeren Raum zu atmen.**
>
> Albert Einstein, 1936 in *Physik und Realität*.

In this rapid journey, we have destroyed all the experimental pillars of quantum theory: the superposition principle, space-time symmetry, gauge symmetry, renormalization symmetry and permutation symmetry. We also have destroyed the foundations of special and general relativity, namely the concepts of the space-time manifold, fields, particles and mass. We have even seen that matter and vacuum cannot be distinguished.

It seems that we have lost every concept used for the description of motion, and thus made its description impossible. It seems that we have completely destroyed our two 'castles in the air', general relativity and quantum theory. And it seems that we are trying to breathe in airless space. Is this pessimistic view correct, or can we save the situation?

First of all, since matter and radiation are not distinguishable from vacuum, the quest for unification in the description of elementary particles is correct and necessary. There is no alternative to tearing down the castles and to continuing to breathe.

---

* 'I consider it as quite possible that physics cannot be based on the field concept, i.e., on continuous structures. In that case, nothing remains of my castle in the air, gravitation theory included.'

** 'Yet it has been suggested that the introduction of a space-time continuum, in view of the molecular structure of all events in the small, may possibly be considered as contrary to nature. Perhaps the success of Heisenberg's method may point to a purely algebraic method of description of nature, to the elimination of continuous functions from physics. Then, however, one must also give up, in principle, the use of the space-time continuum. It is not inconceivable that human ingenuity will some day find methods that will make it possible to proceed along this path. Meanwhile, however, this project resembles the attempt to breathe in an airless space.'

Secondly, after tearing down the castles, one result remains. Since the concepts of 'mass', 'time' and 'space' cannot be distinguished from each other, a new, *single* entity or concept is necessary to define both particles and space-time. In short, vacuum and particles must be made of *common constituents*. In other words, we are not in airless space, and we uncovered the foundations that remain after we tore down the castles. Before we go on exploring these common constituents, we check what we have deduced so far against experiment.

SOME EXPERIMENTAL PREDICTIONS

A race is going on both in experimental and in theoretical physics: to be the first to Challenge 44 r suggest and to be the first to perform an experiment that detects a quantum gravity effect, i.e., a Planck scale effect. Here are some proposals.

At Planck scales, space fluctuates. We might think that the fluctuations of space could blur the images of faraway galaxies, or destroy the phase relation between the photons. Ref. 87 However, this effect has been shown to be unmeasurable in all possible cases.

Another idea is to measure the speed of light at different frequencies from faraway light flashes. There are natural flashes, called gamma ray bursts, which have an extremely broad spectrum, from 100 GeV down to visible light at about 1 eV. These flashes often originate at cosmological distances d. Using *short* gamma ray bursts, it is thus possible to test precisely whether the quantum nature of space-time influences the dispersion of Ref. 91, Ref. 90 light signals when they travel across the universe. Planck-scale quantum gravity effects *might* produce a dispersion. Detecting a dispersion would confirm that Lorentz symmetry breaks down at Planck scales.

The difference in arrival time Δt between two photon energies E_1 and E_2 defines a characteristic energy by

$$E_{\text{char}} = \frac{(E_1 - E_2)d}{c\Delta t} \,. \tag{93}$$

This energy value is between $1.4 \cdot 10^{19}$ GeV and over 10^{22} GeV for the best measurement Ref. 88, Ref. 89 to date. This is between just above the Planck energy and over one thousand times the Ref. 89 Planck energy. However, despite this high characteristic energy, *no dispersion* has been found: even after a trip of ten thousand million years, all light arrives within one or two seconds.

Another candidate experiment is the direct detection of distance fluctuations between Ref. 91, Ref. 92 bodies. Gravitational wave detectors are sensitive to extremely small noise signals in length measurements. There should be a noise signal due to the distance fluctuations induced near Planck energy. The indeterminacy in measurement of a length l is predicted Ref. 93 to be

$$\frac{\delta l}{l} \geqslant \left(\frac{l_{\text{Pl}}}{l}\right)^{2/3} \,. \tag{94}$$

This expression is deduced simply by combining the measurement limit of a ruler, from Page 61 quantum theory, with the requirement that the ruler not be a black hole. The sensitivity of the detectors to noise might reach the required level later in the twenty-first century. The Ref. 94 noise induced by quantum gravity effects has also been predicted to lead to detectable quantum decoherence and vacuum fluctuations. However, no such effect has been found yet.

A further candidate experiment for measuring quantum gravity effects is the detection of the loss of CPT symmetry at high energies. Especially in the case of the decay of certain elementary particles, such as neutral kaons, the precision of experimental measurement is approaching the detection of Planck-scale effects. However, no such effect has been found yet. Ref. 91

Another possibility is that quantum gravity effects may change the threshold energy at which certain particle reactions become possible. It may be that extremely high-energy photons or cosmic rays will make it possible to prove that Lorentz invariance is indeed broken near the Planck scale. However, no such effect has been found yet. Ref. 95

The next possibility is that the phase of light that travels over long distances might get washed out. However, the first tests seem to show that this is not the case: light from extremely distant galaxies still interferes. The precise prediction of the phase washing effect is still being worked out; most probably, the effect is too small to be measured. Ref. 87

In the domain of atomic physics, it has also been predicted that quantum gravity effects will induce a gravitational Stark effect and a gravitational Lamb shift in atomic transitions. However, no such effect has been found yet. Ref. 94

Other proposals start from the recognition that the bound on the measurability of observables also puts a bound on the measurement *precision* for each observable. This bound is of no importance in everyday life, but it is important at Planck energy. One proposal is to search for a minimal noise in length measurements, e.g., in gravitational wave detectors. But no such noise has been found yet.

Another proposal asks about the precision with which a coupling constant can be measured. Let us take the electromagnetic coupling constant α, usually called the fine structure constant. It is related to the positron charge q by

$$q = \sqrt{4\pi\varepsilon_0 \hbar c \alpha} \ . \tag{95}$$

Now, any electric charge is defined and measured by comparing, in an electric field, the acceleration to which the charged object is subjected with the acceleration of some unit charge q_{unit}. In other words, we have

$$\frac{q}{q_{\text{unit}}} = \frac{ma}{m_{\text{unit}} a_{\text{unit}}} \ . \tag{96}$$

Therefore any error in mass and acceleration measurements implies errors in measurements of charge and the coupling constant.

Vol. V, page 89 We found in the part on quantum theory that the electromagnetic, weak and strong interactions are characterized by coupling constants, the inverses of which depend linearly on the logarithm of the energy. We also know from the above discussions that the minimum error for any energy measurement at high energies is given by the ratio between the energy to be measured and the limit energy. If we plot a graph of the coupling constants against energy, we get the result shown in Figure 6. The search for consequences of this *fan-out effect* is still open. One way to express the result is to say that coupling

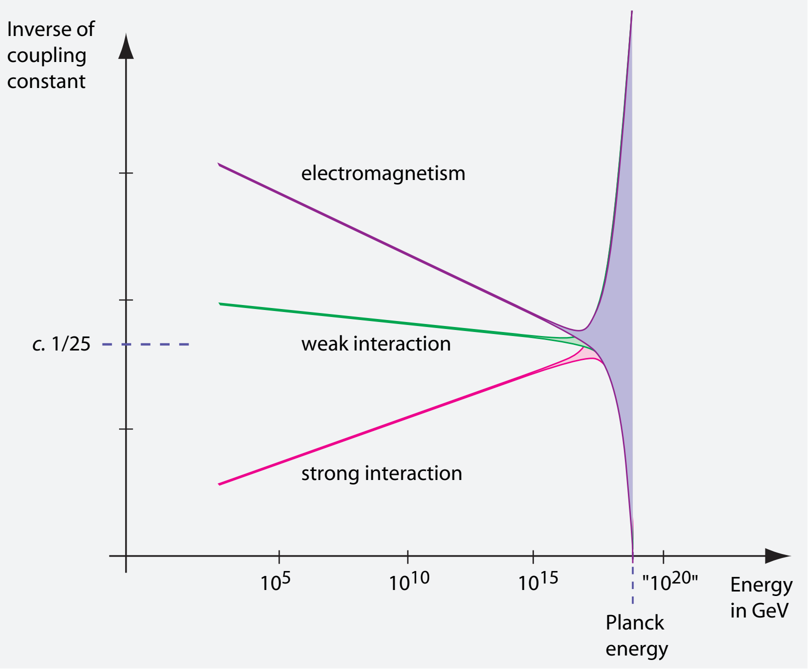


FIGURE 6 Coupling constants and their predicted spread as a function of energy.

constants are by definition subject to an error.

All measurement devices, be they clocks, metre rules, scales or any other devices, use electromagnetic effects at energies of around 1 eV plus the electron rest energy. This is about 10^{-22} times the Planck energy. As a consequence, the measurement precision of any observable should be limited to about 22 digits. The maximum precision currently achieved is 14 digits, and, for the electromagnetic coupling constant, about 9 digits. It will thus be quite some time before this prediction can be tested.

It may also be that high-precision measurements of the g-factor of elementary particles or debris of high-energy cosmic ray reactions will show some effects of the fan-out. The lifetimes of elementary particles could also be affected. Can you find another effect?

Challenge 45 r

In summary, the experimental detection of quantum gravity effects *might* be possible, despite their weakness, at some time during the twenty-first century. The successful detection of such an effect would be one of the highlights of physics, as it would challenge the usual description of space and time even more than general relativity did. On the other hand, most unified models of physics predict the *absence* of any measurable quantum gravity effects. Ref. 96

Summary on particles and vacuum

In summary, combining quantum theory and general relativity leads us to two main results. First, the constituents of vacuum and particles *cannot be points*. There is no conceivable way to prove that points exist, as the smallest measurable distance in nature is the Planck length.

Ref. 113

Secondly, *vacuum and particles mix* at Planck scales, as there is no conceivable way to distinguish whether a Planck-sized region is part of a particle or of empty space. Matter, radiation and vacuum cannot be distinguished at Planck scales. In other words, empty space and particles are made of *common constituents*. This result must be part of the final theory that we are looking for.

Page 51

Generally speaking, we thus found the same conclusions that we found in the chapter on limit statements. We therefore continue in the same way we did there: we explore the universe as a whole.

CHAPTER 5

WHAT IS THE DIFFERENCE BETWEEN THE UNIVERSE AND NOTHING?

> Die Grenze ist der eigentlich fruchtbare Ort der Erkenntnis.*
>
> Paul Tillich, *Auf der Grenze.*

This strange question is the topic of the current leg of our mountain ascent. In he last section we explored nature in the vicinity of Planck scales; but he other limit, namely the description of motion at large, cosmological scales, is equally fascinating. As we proceed, many incredible results will appear, and at the end we will discover a surprising answer to the question in the title.

COSMOLOGICAL SCALES

> Hic sunt leones.**
>
> Antiquity

The description of motion requires the application of general relativity whenever the scale d of the situation is of the order of the Schwarzschild radius, i.e., whenever

$$d \approx r_S = 2Gm/c^2 \ . \tag{97}$$

Challenge 46 s

It is straightforward to confirm that, with the usually quoted mass m and size d of everything visible in the universe, this condition is indeed fulfilled. We do need general relativity, and thus curved space-time, when talking about the whole of nature.

Similarly, quantum theory is required for the description of the motion of an object whenever we approach it within a distance d of the order of the Compton wavelength λ_C, i.e., whenever

$$d \approx \lambda_C = \frac{h}{mc} \ . \tag{98}$$

Obviously, for the total mass of the universe this condition is not fulfilled. However, we are not interested in the motion of the universe itself; we are interested in the motion of its components. In the description of these components, quantum theory is required whenever pair production and annihilation play a role. This is the case in the early his-

---

* 'The frontier is the really productive place of understanding.' Paul Tillich (1886–1965), German theologian, socialist and philosopher.

** 'Here are lions.' Written across unknown and dangerous regions on ancient maps.

tory of the universe and near the horizon, i.e., for the most distant events that we can observe in space and time. We are thus obliged to include quantum theory in any precise description of the universe.

Since at cosmological scales we need both quantum theory and general relativity, we start our investigation with the study of time, space and mass, by asking at large scales the same questions that we asked above at Planck scales.

MAXIMUM TIME

Is it possible to measure time intervals of any imaginable size? General relativity shows that in nature there is a maximum time interval, with a value of about fourteen thousand million years, or 430 Ps, providing an upper limit to the measurement of time. It is called the *age of the universe*, and has been deduced from two sets of measurements: the expansion of space-time and the age of matter.

We are all familiar with clocks that have been ticking for a long time: the hydrogen atoms in our body. All hydrogen atoms were formed just after the big bang. We can almost say that the electrons in these atoms have been orbiting their nuclei since the dawn of time. In fact, the quarks inside the protons in these atoms have been moving a few hundred thousand years longer than the electrons.

Challenge 47 s

We thus have an upper time limit for any clock made of atoms. Even 'clocks' made of radiation (can you describe one?) yield a similar maximum time. The study of the spatial expansion of the universe leads to the *same* maximum age. No clock or measurement device was ticking *longer ago* than this maximum time, and no clock could provide a record of having done so.

In summary, it is not possible to measure time intervals greater than the maximum one, either by using the history of space-time or by using the history of matter or radiation.* The maximum time is thus rightly called the 'age' of the universe. Of course, this is not a new idea; but looking at the issue in more detail does reveal some surprises.

DOES THE UNIVERSE HAVE A DEFINITE AGE?

> One should never trust a woman who tells one
> her real age. A woman who would tell one that,
> would tell one anything.
>
> Oscar Wilde

Vol. II, page 275

In light of all measurements, it may seem silly to question the age of the universe. The age value is found in many books and tables and its precise determination is one of the most important quests in modern astrophysics. But is this quest reasonable?,

In order to measure the duration of a movement or the age of a system, we need a clock that is *independent* of that movement or system, and thus *outside* the system. However, there are no clocks outside the universe, and no clock inside it can be independent. In fact we have just seen that no clock inside the universe can run throughout its full history, in particular, through its earliest history.

---

* This implies that so-called 'oscillating universe' models, in which it is claimed that 'before' the big bang there were other phenomena, cannot be justified on the basis of nature or observations. They are based on beliefs.

Time can be defined only if it is possible to distinguish between matter and space. Given this distinction, we can talk either about the age of *space*, by assuming that matter provides suitable and independent clocks – as is done in general relativity – or about the age of *matter*, such as stars or galaxies, by assuming that the extension of space-time, or some other matter, provides a good clock. Both possibilities are being explored experimentally in modern astrophysics – and both give the same result, of about fourteen thousand million years, which was mentioned above. However, for the universe as a *whole*, an age *cannot* be defined.

The issue of the starting point of time makes this difficulty even more apparent. We may imagine that going back in time leads to only two possibilities: either the starting instant $t = 0$ is part of time or it is not. (Mathematically, this means that the segment representing time is either closed or open.) Both these possibilities imply that it is possible to measure arbitrarily small times; but we know from the combination of general relativity and quantum theory that this is *not* the case. In other words, neither possibility is correct: the beginning cannot *be* part of time, nor can it *not be* part of it. There is only one solution to this contradiction: there was no beginning at all.

Indeed, a minimum length, or equivalently, a minimum action, both imply that there is a maximum curvature for space-time. Curvature can be measured in several ways: for example, surface curvature is an inverse area. Within a factor of order one, we find

$$K < \frac{c^3}{G\hbar} = 0.39 \cdot 10^{70}\,\mathrm{m}^{-2} \tag{99}$$

as a limit for the surface curvature K in nature. In other words, the universe has never been a point, never had zero age, never had infinite density, and never had infinite curvature. It is not difficult to get a similar limit for temperature or any other physical quantity near the big bang. In short, since events do not exist, the big bang cannot have been an event. There never was an initial singularity or a beginning of the universe. Challenge 48 s

In summary, the situation is consistently muddled. Neither the age of the universe nor its origin makes sense. What is going wrong? Or rather, *how* are things going wrong? What happens if instead of jumping directly to the big bang, we *approach* it as closely as possible? The best way to clarify the issue is to ask about the measurement *error* in our statement that the universe is fourteen thousand million years old. This turns out to be a fascinating topic.

HOW PRECISE CAN AGE MEASUREMENTS BE?

> “No woman should ever be quite accurate about her age. It looks so calculating.”
>
> Oscar Wilde

The first way to measure the age of the universe* is to look at clocks in the usual sense of the word, namely at clocks made of *matter*. As explained in the part on quantum theory, Salecker and Wigner showed that a clock built to measure a total time T with a precision Δt has a minimum mass m given by Ref. 54, Ref. 55

$$m > \frac{\hbar}{c^2} \frac{T}{(\Delta t)^2} . \tag{100}$$

A simple way to incorporate general relativity into this result was suggested by Ng and van Dam. Any clock of mass m has a minimum resolution Δt due to the curvature of space that it introduces, given by Ref. 93

$$\Delta t > \frac{Gm}{c^3} . \tag{101}$$

If m is eliminated, these two results imply that a clock with a precision Δt can only measure times T up to a certain maximum value, namely

$$T < \frac{(\Delta t)^3}{t_{\mathrm{Pl}}^2} , \tag{102}$$

where $t_{\mathrm{Pl}} = \sqrt{\hbar G/c^5} = 5.4 \cdot 10^{-44}$ s is the Planck time. (As usual, we have omitted factors of order one in this and in all the following results of this chapter.) In other words, the higher the accuracy of a clock, the shorter the time during which it works dependably. The precision of a clock is limited not only by the expense of building it, but also by nature itself. Nevertheless, it is easy to check that for clocks used in daily life, this limit is not even remotely approached. For example, you may wish to calculate how precisely your own age can be specified. Challenge 49 e

As a consequence of the inequality (102), a clock trying to achieve an accuracy of one Planck time can do so for at most one Planck time! *A real clock cannot achieve Planck-time accuracy.* If we try to go beyond the limit (102), fluctuations of space-time hinder the working of the clock and prevent higher precision. With every Planck time that passes, the clock accumulates a measurement error of at least one Planck time. Thus, the total measurement error is at least as large as the measurement itself. This conclusion is also valid for clocks based on radiation, for example those based on the background radiation. Challenge 50 ny

In short, measuring age with a clock always involves errors. Whenever we try to reduce these errors to the smallest possible level, the Planck level, the clock becomes so imprecise over large times that age measurements become impossible.

---

* Note that the age t_0 is not the same as the Hubble time $T = 1/H_0$. The Hubble time is only a computed quantity and (almost) always larger than the age; the relation between the two depends on the values of the cosmological constant, the density and other parameters of the universe. For example, for the standard 'hot big bang' scenario, i.e., for the matter-dominated Einstein–de Sitter model, we have the simple relation $T = (3/2)\, t_0$. Ref. 97

DOES TIME EXIST?

> Time is waste of money.
>
> Oscar Wilde

Vol. I, page 38

Ever since people began to study physics, the concept of 'time' has designated what is measured by a clock. But the inequality (102) for a maximum clock time implies that perfect clocks do not exist, and thus that time is only an approximate concept: perfect time does not exist. Thus, in nature there is no 'idea' of time, in the Platonic sense. In fact, the discussion so far can be seen as proof that combining quantum theory and general relativity, because of the resulting measurement errors, prevents the existence of perfect or 'ideal' examples of any classical or everyday concept.

Challenge 51 e

Time does not exist. Yet it is obviously a useful concept in everyday life. The key to understanding this is *measurement energy*. Any clock – in fact, any system of nature – is characterized by a simple number, namely the highest ratio of its kinetic energy to the rest energy of its components. In daily life, this ratio is about $1\,\mathrm{eV}/10\,\mathrm{GeV} = 10^{-10}$. Such *low-energy* systems are well suited for building clocks. The more precisely the motion of the main moving part – the pointer of the clock – can be kept constant and monitored, the higher the precision of the clock. To achieve very high precision, the pointer must have very high mass. Indeed, in any clock, both the position and the speed of the pointer must be measured, and the two measurement errors are related by the quantum-mechanical indeterminacy relation $\Delta v\,\Delta x > \hbar/m$. High mass implies low intrinsic fluctuation. Furthermore, in order to screen the pointer from outside influences, even more mass is needed. This connection between mass and accuracy explains why more accurate clocks are usually more expensive.

Page 54
Page 63
Challenge 52 e

The standard indeterminacy relation $m\Delta v\,\Delta x > \hbar$ is valid only at everyday energies. However, we cannot achieve ever higher precision simply by increasing the mass without limit, because general relativity changes the indeterminacy relation to $\Delta v\,\Delta x > \hbar/m + G(\Delta v)^2 m/c^3$. The additional term on the right-hand side, negligible at everyday scales, is proportional to energy. Increasing it by a large amount limits the achievable precision of the clock. The smallest measurable time interval turns out to be the Planck time.

In summary, time exists, as a good approximation, only for *low-energy* systems. Any increase in precision beyond a certain limit requires an increase in the energy of the components; at Planck energy, this increase will prevent an increase in precision.

WHAT IS THE ERROR IN THE MEASUREMENT OF THE AGE OF THE UNIVERSE?

Challenge 53 e

It is now straightforward to apply our discussion about the measurement of time to the age of the universe. The inequality (102) implies that the highest precision possible for a clock is about 10^{-23} s, or about the time light takes to move across a proton. The finite age of the universe also yields a maximum *relative* measurement precision. Inequality (102) can be written as

$$\frac{\Delta t}{T} > \left(\frac{t_{\mathrm{Pl}}}{T}\right)^{2/3} . \qquad (103)$$

Inserting the age of the universe for T, we find that no time interval can be measured with a precision of more than about 40 decimals.

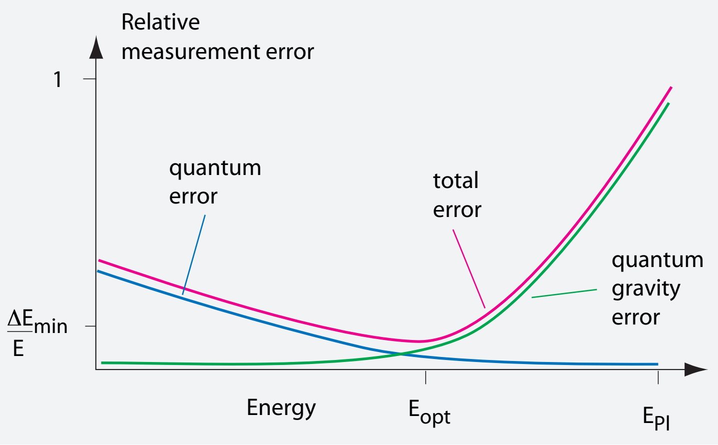


FIGURE 7 Measurement errors as a function of measurement energy.

To clarify the issue, we can calculate the error in measurement as a function of the observation energy E_{meas}, the energy of the measurement probe. There are two limit cases. For *low* energies, the error is due to quantum effects and is given by

$$\frac{\Delta t}{T} \sim \frac{1}{E_{\text{meas}}} \tag{104}$$

which decreases with increasing measurement energy. For *high* energies, however, the error is due to gravitational effects and is given by

$$\frac{\Delta t}{T} \sim \frac{E_{\text{meas}}}{E_{\text{Pl}}} \tag{105}$$

so that the total error varies as shown in Figure 7. In particular, very high energies do not reduce measurement errors: any attempt to reduce the measurement error for the age of the universe below 10^{-23} s would require energies so high that the limits of space-time would be reached, making the measurement itself impossible. We reached this conclusion through an argument based on clocks made of particles. We will see below that trying to determine the age of the universe from its expansion leads to the same limitation.

Imagine observing a tree which, as a result of some storm or strong wind, has fallen towards second tree, touching it at the very top, as shown in Figure 8. It is possible to determine the heights of both trees by measuring their separation and the angles at the base. The *error* in the heights will depend on the errors in measurement of the separation and angles.

Similarly, the age of the universe can be calculated from the present distance and speed of objects – such as galaxies – observed in the night sky. The present distance d corresponds to separation of the trees at ground level, and the speed v to the angle between the two trees. The Hubble time T of the universe (which is usually assumed to be larger than the age of the universe) then corresponds to the height at which the two

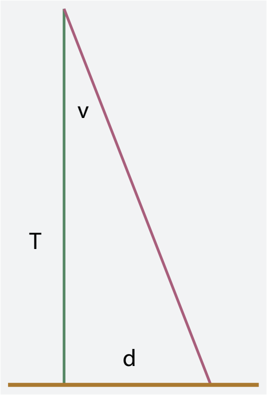


FIGURE 8 Trees and galaxies.

trees meet. This age – in a naive sense, the time since the galaxies 'separated' – is given, within a factor of order one, by

$$T = \frac{d}{v} \; . \tag{106}$$

In simple terms, this is the method used to determine the age of the universe from the expansion of space-time, for galaxies with red-shifts below unity.* The (positive) measurement error ΔT becomes

$$\frac{\Delta T}{T} = \frac{\Delta d}{d} + \frac{\Delta v}{v} \; . \tag{107}$$

It is worthwhile to explore this in more detail. For any measurement of T, we have to choose the object, i.e., a distance d, as well as an observation time Δt, or, equivalently, an observation energy $\Delta E = 2\pi\hbar/\Delta t$. We will now investigate the consequences of these choices for equation (107), always taking into account both quantum theory and general relativity.

At everyday energies, the result of the determination of the age of the universe t_0 is about $(13.7 \pm 0.2) \cdot 10^9$ Ga. This value is deduced by measuring red-shifts, i.e., velocities, and distances, using stars and galaxies in distance ranges, from some hundred thousand light years up to a red-shift of about 1. Measuring red-shifts does not produce large velocity errors. The main source of experimental error is the difficulty in determining the distances of galaxies.

What is the smallest possible error in distance? Obviously, inequality (103) implies

$$\frac{\Delta d}{T} > \left(\frac{l_{\text{Pl}}}{d} \right)^{2/3} \tag{108}$$

thus giving the same indeterminacy in the age of the universe as the one we found above in the case of material clocks.

Challenge 54 e

---

* At higher red-shifts, the speed of light, as well as the details of the expansion, come into play. To continue with the analogy of the trees, we find that the trees are not straight all the way up to the top and that they grow on a slope, as suggested by Figure 9.

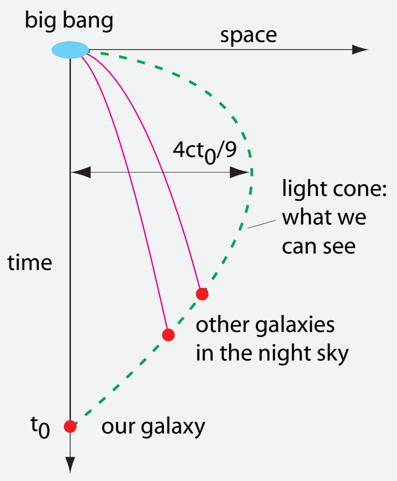


FIGURE 9 The speed and distance of remote galaxies.

Challenge 55 e

We can try to reduce the age error in two ways: by choosing objects at either small or large distances. Let us start with small distances. In order to get high precision at small distances, we need high observation energies. It is fairly obvious that at observation energies near the Planck value, $\Delta T/T$ approaches unity. In fact, both terms on the right-hand side of equation (107) become of order one. At these energies, Δv approaches c and the maximum value for d approaches the Planck length, for the same reason that at Planck energy the maximum measurable time is the Planck time. In short, *at Planck scales it is impossible to say whether the universe is old or young.*

Ref. 97

Let us consider the other extreme, namely objects extremely far away, say with a red-shift of $z \gg 1$. Relativistic cosmology requires the diagram of Figure 8 to be replaced by the more realistic diagram of Figure 9. The 'light onion' replaces the familiar light cone of special relativity: light converges near the big bang. In this case the measurement error for the age of the universe also depends on the distance and velocity errors. At the largest possible distances, the signals an object sends out must be of high energy, because the emitted wavelength must be smaller than the universe itself. Thus, inevitably, we reach Planck energy. However, we have seen that in such high-energy situations, the emitted radiation, as well as the object itself, is indistinguishable from the space-time background. In other words, the red-shifted signal we would observe today would have a wavelength as large as the size of the universe, with a correspondingly small frequency.

There is another way to describe the situation. At Planck energy or near the horizon, the original signal has an error of the same size as the signal itself. When measured at the present time, the red-shifted signal still has an error of the same size as the signal. As a result, the error in the horizon distance becomes as large as the value to be measured.

In short, even if space-time expansion and large scales are used, the instant of the so-called beginning of the universe cannot be determined with an error smaller than the age of the universe itself: a result we also found at Planck distances. If we aim for perfect precision, we just find that the universe is 13.7 ± 13.7 thousand million years old! In other words, *in both extremal situations, it is impossible to say whether the universe has a non-vanishing age.*

We have to conclude that the anthropocentric concept of 'age' does not make any sense for the universe as a whole. The usual textbook value is useful only for ranges of time, space and energy in which matter and space-time are clearly distinguished, namely at everyday, human-scale energies; the value has no more general meaning.

Challenge 56 ny

You may like to examine the issue of the *fate* of the universe using the same arguments. But we will now continue on the path outlined at the start of this chapter: the next topic is the measurement of length.

MAXIMUM LENGTH

Vol. II, page 217

General relativity shows that the *horizon distance*, i.e., the distance of objects with infinite red-shift, is finite. In the usual cosmological model, for hyperbolic (open) and parabolic (marginal) evolutions of the universe, the *size* of the universe is assumed infinite.* For elliptical evolution, the total size is finite and depends on the curvature. However, in this case also the present measurement limit yields a minimum size for the universe many times larger than the horizon distance.

Quantum field theory, on the other hand, is based on flat and infinite space-time. Let us see what happens when the two theories are combined. What can we say about measurements of length in this case? For example, would it be possible to construct and use a metre rule to measure lengths larger than the distance to the horizon?

Ref. 97

Ref. 97

Admittedly, we would have no time to push the metre rule out up to the horizon, because in the standard Einstein–de Sitter big bang model the horizon moves away from us faster than the speed of light. (We should have started using the metre rule right at the big bang.) But just for fun, let us assume that we have actually managed to do this. How far away can we read off distances? In fact, since the universe was smaller in the past, and since every observation of the sky is an observation of the past, Figure 9 shows that the maximum *spatial distance* away from us at which an object can be seen is only $4ct_0/9$. Obviously, for space-time intervals, the maximum remains ct_0.

Thus, in all cases it turns out to be impossible to measure lengths larger than the horizon distance, even though general relativity sometimes predicts such larger distances. This result is unsurprising, and in obvious agreement with the existence of a limit for measurements of time intervals. The real surprises come next.

IS THE UNIVERSE REALLY A BIG PLACE?

Ref. 98

Vol. II, page 275

Astronomers and Hollywood films answer this question in the affirmative. Indeed, the distance to the horizon of the universe is often included in tables. Cosmological models specify that the scale factor R, which fixes the distance to the horizon, grows with time t; for the case of the standard mass-dominated Einstein–de Sitter model, i.e., for a vanishing cosmological constant and flat space, we have

$$R(t) = C\ t^{2/3}\ , \qquad (109)$$

Ref. 97

* In cosmology, we need to distinguish between the scale factor R, the Hubble radius $c/H = cR/\dot{R}$, the horizon distance h and the size d of the universe. The Hubble radius is a computed quantity giving the distance at which objects move away with the speed of light. The Hubble radius is always *smaller* than the horizon distance, at which in the standard Einstein–de Sitter model, for example, objects move away with *twice* the speed of light. However, the horizon itself moves away with *three* times the speed of light.

where the numerical constant C relates the commonly accepted horizon distance to the commonly accepted age. Indeed, observation shows that the universe is large, and is getting larger. But let us investigate what happens if we add some quantum theory to this result from general relativity. Is it really possible to measure the distance to the horizon?

We look first at the situation at high (probe) energies. We saw above that space-time and matter are not distinguishable at Planck scales. Therefore, at Planck energy we cannot state whether or not objects are *localized*. At Planck scales, the distinction between matter and vacuum – so basic to our thinking – disappears.

Another way to say this is that we cannot claim that space-time is *extended* at Planck scales. Our concept of extension derives from the possibility of measuring distances and time intervals, and from observations such as the ability to align several objects behind one another. Such observations are not possible at Planck scales and energies, because of the inability of probes to yield useful results. In fact, all of the everyday observations from which we deduce that space is extended are impossible at Planck scales, where *the basic distinction between vacuum and matter, namely between extension and localization, disappears*. As a consequence, at Planck energy the size of the universe cannot be measured. It cannot even be called larger than a matchbox.

The problems encountered with probes of high probe energies have drastic consequences for the size measurement of the universe. All the arguments given above for the errors in measurement of the age can be repeated for the distance to the horizon. To reduce size measurement errors, a measurement probe needs to have high energy. But at high energy, measurement errors approach the value of the measurement results. At the largest distances and at Planck energy, the measurement errors are of the same magnitude as the measured values. If we try to determine the size of the universe with high precision, we get no precision at all.

Challenge 57 ny

The inability to get precise values for the size of the universe should not come unexpected. For a reliable measurement, the standard must be different, independent, and outside the system to be measured. For the universe this is impossible.

Studying the size of the big bang also produces strange results. The universe is said to have been much smaller near the big bang because, on average, all matter is moving away from all other matter. But if we try to follow the path of matter into the past with high precision, using Planck energy probes, we get into trouble: since measurement errors are as large as measurement data, we cannot claim that the universe was smaller near the big bang than it is today: there is no way to reliably distinguish size values.

There are other confirmations too. If we had a metre rule spanning the whole universe, even beyond the horizon, with zero at the place where we live, what measurement *error* would it produce for the horizon? It does not take long to work out that the expansion of space-time, from Planck scales to the present size, implies an expansion in the error from Planck size to a length of the order of the present distance to the horizon. Again, the error is as large as the measurement result. Again, the size of the universe turns out not to be a meaningful property.

Challenge 58 ny

Since this reasoning also applies if we try to measure the diameter of the universe instead of its radius, it is impossible to say whether the antipodes in the sky really are distant from each other!

We can summarize the situation by noting that anything said about the size of the universe is as limited as anything said about its age. *The height of the sky depends on* *the observation energy.* If we start measuring the sky at standard observation energies, and try to increase the precision of measurement of the distance to the horizon, the measurement error increases beyond all bounds. At Planck energy, the volume of the universe is indistinguishable from the Planck volume – and vice versa.

The boundary of space – is the sky a surface?

The horizon of the universe – essentially, the black part of the night sky – is a fascinating entity. Everybody interested in cosmology wants to know what happens there. In newspapers the horizon is sometimes called the *boundary of space*. Some surprising insights – which have not yet made it to the newspapers – appear when we combine general relativity and quantum mechanics.

We saw above that the errors in measuring the distance of the horizon are substantial. They imply that we cannot pretend that all points of the sky are equally far away from us. Thus we cannot say that the sky is a surface; it could be a volume. In fact, there is no way to determine the dimensionality of the horizon, or the dimensionality of space-time near the horizon.*

Thus measurements do not allow us to determine whether the boundary is a point, a surface, or a line. It may be a very complex shape, even knotted. In fact, quantum theory tells us that it must be all of these from time to time: that *the sky fluctuates in height and shape*. In short, it is impossible to determine the topology of the sky. But that is nothing new. As is well known, general relativity is unable to describe pair creation particles with spin 1/2. The reason for this is the change in space-time topology required by the process. The universe is full of such processes, implying that it is impossible to define a topology for the universe and, in particular, for the horizon. Can you find at least two other arguments to show this? Challenge 60 ny

Worse still, quantum theory shows that space-time is not continuous at a horizon: this can easily be deduced using the Planck-scale arguments from the previous section. Time and space are not defined there. Page 54

Finally, there is no way to decide whether the various boundary points are *different* from each other. The distance between two points in the night sky is undefined. Therefore it is unclear what the *diameter* of the horizon is.

In summary, the horizon has no specific distance or shape. The horizon, and thus the universe, cannot be shown to be manifolds. This leads us to the next question.

Does the universe have initial conditions?

Ref. 99 One often reads about the quest for the initial conditions of the universe. But before joining this search, we should ask *whether* and *when* such initial conditions make any sense.

Obviously, our everyday description of motion requires knowledge of initial conditions, which describe the *state* of a system, i.e., all those aspects that differentiate it from

---

* The measurement errors also imply that we cannot say anything about translational symmetry at cosmological scales. Can you confirm this? Challenge 59 ny In addition, at the horizon it is impossible to distinguish between spacelike and timelike distances. Even worse, concepts such as 'mass' or 'momentum' become muddled at the horizon. This means that, as at Planck energy, we are unable to distinguish between object and background, and between state and intrinsic properties. We will come back to this important point shortly.

a system with the same intrinsic properties. Initial conditions – like the state – are attributed to a system by an *outside* observer.

Quantum theory tells us that initial conditions, or the state of a system, can only be defined by an outside observer with respect to an environment. It is already difficult to be outside the universe – but even inside the universe, a state can only be defined if matter can be distinguished from vacuum. This is impossible at Planck energy, near the big bang, or at the horizon. Thus the universe has no state. This means also that it has *no wave function*.

Page 54 The limits imposed by the Planck values confirm this conclusion in other ways. First of all, they show that the big bang was not a singularity with infinite curvature, density or temperature, because infinitely large values do not exist in nature. Secondly, since instants of time do not exist, it is impossible to define the state of any system at a given time. Thirdly, as instants of time do not exist, neither do events, and so the big bang was not an event, and neither an initial state nor an initial wave function can be ascribed to the universe. (Note that this also means that the universe cannot have been created.)

In short, *there are no initial conditions for the universe*. Initial conditions make sense only for subsystems, and only far from Planck scales. Thus, for initial conditions to exist, the system must be far from the horizon and it must have evolved for some time 'after' the big bang. Only when these two requirements are fulfilled can objects *move* in space. Of course, this is always the case in everyday life.

At this point in our mountain ascent, where neither time nor length is clearly defined at cosmological scales, it should come as no surprise that there are similar difficulties concerning the concept of mass.

Does the universe contain particles and stars?

The total number of stars in the universe, about $10^{23\pm1}$, is included in every book on cosmology. A smaller number can be counted on clear nights. But how dependable is the statement? Vol. II, page 275

We can ask the same question about particles instead of stars. The commonly quoted numbers are $10^{80\pm1}$ baryons and $10^{89\pm1}$ photons. However, the issue is not simple. Neither quantum theory nor general relativity alone make predictions about the number of particles, either inside or outside the horizon. What happens if we combine the two theories?

In order to define the number of particles in a region, quantum theory first of all requires a vacuum state to be defined. The number of particles is defined by comparing the system with the vacuum. If we neglect or omit general relativity by assuming flat space-time, this procedure poses no problem. However, if we include general relativity, and thus a curved space-time, especially one with a strangely behaved horizon, the answer is simple: there is *no* vacuum state with which we can compare the universe, for two reasons. First, nobody can explain what an empty universe would look like. Secondly, and more importantly, there is no way to define a state of the universe. The number of particles in the universe thus becomes undefinable. Only at everyday energies and for finite dimensions are we able to speak of an approximate number of particles. Vol. IV, page 97

Comparison between a system and the vacuum is also impossible, in the case of the universe, for purely practical reasons. The particle counter would have to be outside the

Challenge 61 ny system. (Can you confirm this?) In addition, it is impossible to remove particles from the universe.

The impossibility of defining a vacuum state, and thus the number of particles in the universe, is not surprising. It is an interesting exercise to investigate the measurement errors that appear when we try to determine the number of particles despite this fundamental impossibility. Challenge 62 ny

Can we count the stars? In principle, the same conclusion applies as for particles. However, at everyday energies the stars can be counted *classically*, i.e., without taking them out of the volume in which they are enclosed. For example, this is possible if the stars are differentiated by mass, colour or any other individual property. Only near Planck energy or near the horizon are these methods inapplicable. In short, the number of stars is only defined as long as the observation energy is low, i.e., as long as we stay away from Planck energy and from the horizon.

So, despite appearances on human scales, *there is no definite number of particles in the universe*. The universe cannot be distinguished from vacuum by counting particles. Even though particles are necessary for our own existence and functioning, a complete count of them cannot be made.

This conclusion is so strange that we should try to resist it. Let us try another method of determining the content of matter in the universe: instead of counting particles, let us weigh them.

Does the universe contain masses and objects?

Vol. II, page 275 The average density of the universe, about $10^{-26}\,\mathrm{kg/m^3}$, is often cited in texts. Is it different from a vacuum? Quantum theory shows that, as a result of the indeterminacy relation, even an empty volume of size R has a mass. For a zero-energy photon inside such a vacuum, we have $E/c = \Delta p > \hbar/\Delta x$, so that in a volume of size R, we have a minimum mass of at least $m_{\mathrm{min}}(R) = h/cR$. For a spherical volume of radius R there is thus a minimal mass density given approximately by

$$\rho_{\mathrm{min}} \approx \frac{m_{\mathrm{min}}(R)}{R^3} = \frac{\hbar}{cR^4} . \qquad (110)$$

For the universe, if the standard horizon distance R_0 of 14 000 million light years is inserted, the value becomes about $10^{-142}\,\mathrm{kg/m^3}$. This describes the density of the vacuum. In other words, the universe, with a density of about $10^{-26}\,\mathrm{kg/m^3}$, seems to be clearly different from vacuum. But are we sure?

We have just deduced that the radius of the horizon is undefined: depending on the observation energy, it can be as small as the Planck length. This implies that the density of the universe lies somewhere between the lowest possible value, given by the density of vacuum just mentioned, and the highest possible one, namely the Planck density.* In

---

* In fact, at everyday energies the density of the universe lies midway between the two values, yielding the strange relation Challenge 63 ny

$$m_0^2/R_0^2 \approx m_{\mathrm{Pl}}^2/R_{\mathrm{Pl}}^2 = c^4/G^2 . \qquad (111)$$

Vol. V, page 114 But this is nothing new. The approximate equality can be deduced from equation 16.4.3 (p. 620) of STEVEN WEINBERG, *Gravitation and Cosmology*, Wiley, 1972, namely $G n_b m_p = 1/t_0^2$. The relation is short, the relation (110) does *not* provide a clear statement.

Vol. I, page 88 Vol. II, page 58 Another way to measure the mass of the universe would be to apply the original definition of mass, as given in classical physics and as modified by special relativity. Thus, let us try to collide a standard kilogram with the universe. It is not hard to see that whatever we do, using either low or high energies for the standard kilogram, the mass of the universe cannot be constrained by this method. Challenge 64 ny We would need to produce or to measure a velocity change Δv for the rest of the universe after the collision. To hit all the mass in the universe at the same time, we need high energy; but then we are hindered by Planck energy effects. In addition, a properly performed collision measurement would require a mass outside the universe, which is rather difficult to achieve.

Yet another way to measure the mass would be to determine the gravitational mass of the universe through straightforward weighing. But the lack of balances outside the universe makes this an impractical solution, to say the least.

Another way out might be to use the most precise definition of mass provided by general relativity, the so-called *ADM mass*. However, the definition of this requires a specified behaviour at infinity, i.e., a background, which the universe lacks.

We are then left with the other general-relativistic method: determining the mass of the universe by measuring its average curvature. Let us take the defining expressions for Vol. II, page 170 average curvature κ for a region of size R, namely

$$\kappa = \frac{1}{r^2_{\text{curvature}}} = \frac{3}{4\pi}\,\frac{4\pi R^2 - S}{R^4} = \frac{15}{4\pi}\,\frac{4\pi R^3/3 - V}{R^5} \,. \tag{112}$$

We have to insert the horizon radius R_0 and either its surface area S_0 or its volume V_0. However, given the error margins on the radius and the volume, especially at Planck energy, we again find no *reliable* result for the radius of curvature. Challenge 65 ny

Ref. 100 An equivalent method starts with the usual expression provided by Rosenfeld for the indeterminacy $\Delta\kappa$ in the scalar curvature for a region of size R, namely

$$\Delta\kappa > \frac{16\pi l^2_{\text{Pl}}}{R^4} \,. \tag{113}$$

However, this expression also shows that the error in the radius of curvature behaves like the error in the distance to the horizon.

In summary, *at Planck energy, the average radius of curvature of nature lies between infinity and the Planck length.* This implies that the density of matter lies between the minimum value and the Planck value. There is thus no method to determine the mass of the universe at Planck energy. (Can you find one?) Challenge 66 ny The concept of mass cannot be applied to the universe as a whole. Thus, *the universe has no mass.*

DO SYMMETRIES EXIST IN NATURE?

We have already seen that at the horizon, space-time translation symmetry breaks down. Let us have a quick look at the other symmetries.

---

required by several cosmological models.

What happens to permutation symmetry? Permutation is an operation on objects in space-time. It thus necessarily requires a distinction between matter, space and time. If we cannot distinguish positions, we cannot talk about exchange of particles. Therefore, at the horizon, general relativity and quantum theory together make it impossible to define permutation symmetry.

The same is true of CPT symmetry. As a result of measurement errors or of limiting maximum or minimum values, it is impossible to distinguish between the original and the transformed situations. Therefore we cannot claim that CPT is a symmetry of nature at horizon scales. In other words, matter and antimatter cannot be distinguished at the horizon.

Challenge 67 ny The same is true of gauge symmetry, as you may wish to check in detail yourself. For its definition, the concept of gauge field requires a distinction between time, space and mass; at the horizon this is impossible. We therefore also deduce that at the horizon, concepts such as algebras of observables cannot be used to describe nature. Renormalization breaks down too.

All symmetries of nature break down at the horizon. None of the vocabulary we use to talk about observations – including terms such as such as 'magnetic field', 'electric field', 'potential', 'spin', 'charge', or 'speed' – can be used at the horizon.

Does the universe have a boundary?

It is common to take 'boundary' and 'horizon' as synonyms in the case of the universe, because they are the same for all practical purposes. Knowledge of mathematics does not help us here: the properties of mathematical boundaries – for example, that they themselves have no boundary – are not applicable to the universe, since space-time is not continuous. We need other, physical arguments.

The boundary of the universe is supposed to represent the boundary between *something* and *nothing*. There are three possible interpretations of 'nothing':

- 'Nothing' could mean 'no matter'. But we have just seen that this distinction cannot be made at Planck scales. So either the boundary will not exist at all or it will encompass the horizon *as well as* the whole universe.
- 'Nothing' could mean 'no space-time'. We then have to look for those domains where space and time cease to exist. These occur at Planck scales and at the horizon. Again, either the boundary will not exist or it will encompass the whole universe.
- 'Nothing' could mean 'neither space-time nor matter.' The only possibility is a boundary that encloses domains *beyond* the Planck scale and *beyond* the horizon; but again, such a boundary would also encompass all of nature.

This is puzzling. When combining quantum theory and relativity, we do not seem to be able to find a conceptual definition of the horizon that distinguishes it from what it includes. (If you find one, publish it!) A distinction *is* possible in general relativity alone, and in quantum theory alone; but as soon as we combine the two, the boundary becomes indistinguishable from its content. *The interior of the universe cannot be distinguished from its horizon*. There is no boundary. Challenge 68 ny

The difficulty in distinguishing the horizon from its contents suggests that nature may be *symmetric* under transformations that exchange interiors and boundaries. This idea is called *holography*, because it vaguely recalls the working of credit-card holograms. It is a busy research field in high-energy physics. However, for the time being, we shall continue with our original theme, which leads us to our next question. Ref. 101

IS THE UNIVERSE A SET? – AGAIN

> “Domina omnium et regina ratio.*”
>
> Cicero

We are used to thinking of the universe the sum of all matter and all space-time. In doing so, we imply that the universe is a set of mutually distinct components. This idea has been assumed in three situations: in claiming that matter consists of particles; that space-time consists of events (or points); and that different states consist of different initial conditions. However, our discussion shows that the universe is *not* a set of such distinguishable elements. We have encountered several proofs: at the horizon, at the big bang and at Planck scales, it becomes impossible to distinguish between events, between particles, between observables, and between space-time and matter. In those domains, distinctions of any kind become impossible. We have found that distinguishing between two entities – for example, between a toothpick and a mountain – is only *approximately* possible. It is approximately possible because we live at energies well below the Planck energy. The approximation is so good that we do not notice the error when we distinguish cars from people and from toothpicks. Nevertheless, our discussion of the situation at Planck energy shows that a perfect distinction is impossible in principle. *It is impossible to split the universe into separate parts.*

Another way to reach this result is the following. Distinguishing between two entities requires different measurement results: for example, different positions, masses or sizes. Whatever quantity we choose, at Planck energy the distinction becomes impossible. Only at everyday energies is it approximately possible.

In short, since the universe contains no distinguishable parts, there are *no elements* in nature. Simply put: *the universe is not a set*. We envisaged this possibility earlier on; now it is confirmed. The concepts of ‘element’ and ‘set’ are already too specialized to describe the universe. *The universe must be described by a mathematical concept that does not contain any set.* The new concept must be more general than that of a set. Vol. III, page 236

This is a powerful result: a precise description of the universe cannot use any concept that presupposes the existence of sets. But all the concepts we have used so far to describe nature, such as space-time, metric, phase space, Hilbert space and its generalizations, are based on elements and sets. They must all be abandoned at Planck energies, and in any precise description.

Elements and sets must be abandoned. Note that this radical conclusion is deduced from only two statements: the necessity of using quantum theory whenever the dimensions are of the order of the Compton wavelength, and of using general relativity whenever the dimensions are of the order of the Schwarzschild radius. Together, they mean that no precise description of nature can contain elements and sets. The difficulties in complying with this result explain why the unification of the two theories has not so far been successful. Not only does unification require that we stop using space, time

---

* ‘The mistress and queen of all things is reason.’ *Tuscalanae Disputationes*, 2.21.47.

and mass for the description of nature; it also requires that all distinctions, of any kind, should be only approximate. But all physicists have been educated on the basis of exactly the opposite creed!

Many past speculations about the final unified description of nature depend on sets. In particular, all studies of quantum fluctuations, mathematical categories, posets, complex mathematical spaces, computer programs, Turing machines, Gödel's theorem, creation of any sort, space-time lattices, quantum lattices and Bohm's unbroken wholeness presuppose sets. In addition, all speculations by cosmologists about the origin of the universe presuppose sets. But since these speculations presuppose sets, they are wrong. You may also wish to check the religious explanations you know against this criterion. In fact, no approach used by theoretical physicists up to the year 2000 satisfied the requirement that elements and sets must be abandoned.

Ref. 102

Challenge 69 e

The task of abandoning sets is not easy. This is shown with a simple test: do you know of a single concept not based on elements or sets?

Challenge 70 s

The universe is not a set. Therefore, *the universe is not a physical system.* Specifically, it has no state, no intrinsic properties, no wave function, no initial conditions, no density, no entropy and no cosmological constant. The universe is thus neither thermodynamically closed nor open; and it contains no information. All thermodynamic quantities, such as entropy, temperature and free energy, are defined using *ensembles*. Ensembles are limits of systems which are thermodynamically either open or closed. As the universe is neither open nor closed, no thermodynamic quantity can be defined for it.[*] All physical properties are defined only for parts of nature which are approximated or idealized as sets, and thus are physical systems.

CURIOSITIES AND FUN CHALLENGES ABOUT THE UNIVERSE

> Insofern sich die Sätze der Mathematik auf die Wirklichkeit beziehen, sind sie nicht sicher, und sofern sie sicher sind, beziehen sie sich nicht auf die Wirklichkeit.[**]
>
> Albert Einstein

> Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk.[***]
>
> Leopold Kronecker

In mathematics, $2 + 2 = 4$. This statement is an idealization of statements such as 'two apples plus two apples makes four apples.' However, we now know that at Planck energy, the statement about apples is not a correct statement about nature. At Planck energy, objects cannot be counted, because separation of objects is not possible at that scale. We can count objects only because we live at energies much lower than the Planck energy.

---

[*] Some people knew this long before physicists. For example, the belief that the universe is or contains information was ridiculed most thoroughly in the popular science-fiction parody by DOUGLAS ADAMS, *The Hitchhiker's Guide to the Galaxy*, 1979, and its sequels.

[**] 'In so far as mathematical statements describe reality, they are not certain, and as far as they are certain, they are not a description of reality.'

[***] 'God made the integers, all else is the work of man.' Leopold Kronecker (1823–1891) was a well-known mathematician. Among others, the Kronecker delta and the Kronecker product are named for him.

The statement by Kronecker must thus be amended. Since all integers are low-energy approximations, and since we always use low-energy approximations when talking or thinking, we conclude: man also makes the integers.

* *

If vacuum cannot be distinguished from matter or radiation, and if the universe cannot be distinguished from nothing, then it is incorrect to claim that "the universe appeared from nothing." The naive idea of creation is a logical impossibility. "Creation" results from a lack of imagination.

* *

In 2002, Seth Lloyd estimated how much information the universe can contain, and how many calculations it has performed since the big bang. Ref. 103 This estimate is based on two ideas: that the number of particles in the universe is a well-defined quantity, and that the universe is a computer, i.e., a physical system. We now know that neither assumption is correct. This shows the power of the criteria that we have deduced for any precise or complete description of motion.

* *

People take pictures of the cosmic background radiation and its variations. Is it possible that these photographs will show that the spots in one direction of the sky are exactly the same as those in the diametrically opposite direction? Challenge 71 ny

* *

In 1714, Leibniz published his *Monadologie*. In it he explores what he calls a simple substance, which he defined to be a substance that has no parts. He called it a *monad* and describes some of its properties. Ref. 104 However, mainly thanks to his incorrect deductions, the term has not been generally adopted. What is the physical concept most closely related to that of a monad? Challenge 72 s

* *

We usually speak of *the* universe, implying that there is only one of them. Yet there is a simple case to be made that 'universe' is an observer-dependent concept, since the idea of 'all' is observer-dependent. Does this mean that there are many universes, or a 'multiverse'? Challenge 73 s

* *

If all particles were removed (assuming one knew where to put them), there wouldn't be much of a universe left. True? Challenge 74 s

* *

At Planck energy, interactions cannot be defined. Therefore, 'existence' cannot be defined. In short, at Planck energy we cannot say whether particles exist. True? Challenge 75 s

Hilbert's sixth problem settled

Vol. III, page 208 Ref. 105 In the year 1900, David Hilbert gave a famous lecture in which he listed 23 of the great challenges facing mathematics in the twentieth century. Most of these provided challenges to many mathematicians for decades afterwards. A few are still unsolved, among them the sixth, which challenged mathematicians and physicists to find an *axiomatic* treatment of physics. It has remained in the minds of many physicists since that time.

When we combine quantum theory and general relativity, we must abandon the idea of point particle, of space point, and of event. Mathematically speaking, when we combine quantum theory and general relativity, we find that nature does not contain sets, and that the universe is not a set. However, *all* mathematical systems – be they algebraic systems, order systems, topological systems or a mixture of these – are based on elements and sets. Mathematics does not have axiomatic systems that do not contain elements and sets. The reason for this is simple: any (mathematical) concept contains at least one element and one set. However, nature does not. And since nature does not contain sets, an axiomatic description of nature is *impossible*.

All concepts used in physics before the year 2000 depend on elements and sets. For humans, it is difficult even to *think* without first defining a set of possibilities. Yet nature does not contain sets. There is no axiomatic description of nature. And since an axiomatic formulation of physics is impossible, we conclude that the final, unified theory cannot be based on axioms. This is surprising at first, because separate axiomatic treatments of quantum theory and general relativity *are* possible. However, *axiomatic systems in physics are always approximate.* The need to abandon axioms is one of the reasons why reaching a unified description of nature is a challenge.

The impossibility of an axiomatic system for physics is also confirmed in another way. Physics starts with a *circular* definition: space-time is defined with the help of objects and Vol. I, page 334 objects are defined with the help of space-time. In fact, physics has *never* been axiomatic! Physicists have always had to live with circular definitions.

The situation is similar to a child's description of the sky as 'made of air and clouds'. Looking closely, we discover that clouds are made up of water droplets. However, there is air inside clouds, and there is also water vapour away from clouds. When clouds and air are viewed through a microscope, there is no clear boundary between the two. We cannot define either of the terms 'cloud' and 'air' without the other.

Like clouds and air, also objects and vacuum are indistinguishable. Virtual particles are found in vacuum, and vacuum is found inside objects. At Planck scales there is no clear boundary between the two; we cannot define either of the terms 'particle' and 'vacuum' without the other. But despite the lack of a clean definition, and despite the logical problems that can ensue, in both cases the description works well at large, everyday scales.

In summary, an axiomatic description of nature is impossible. The final, unified theory must contain circular definitions. More details on the final description will appear as we continue in our ascent of motion mountain.

The perfect physics book

Since the universe is not a set and since it contains no information, the paradox of the Vol. I, page 333 perfect physics book disappears. A *perfect* physics book describes all of nature. In par-

TABLE 3 Physical statements about the universe when explored at highest precision, i.e., at Planck scales

The universe has no age.	The universe has no beginning.
The universe has no size.	The universe has no volume.
The universe has no shape.	The universe's particle number is undefined.
The universe has no mass.	The universe has no energy.
The universe has no density.	The universe contains no matter.
The universe has no cosmological constant.	The universe has no initial conditions.
The universe has no state.	The universe has no wave function.
The universe is not a physical system.	The universe contains no information.
The universe is not isolated.	The universe is not open.
The universe has no boundaries.	The universe does not interact.
The universe cannot be measured.	The universe cannot be said to exist.
The universe cannot be distinguished from nothing.	The universe cannot be distinguished from a single event.
The universe contains no moments.	The universe is not composite.
The universe is not a set.	The universe is not a concept.
The universe cannot be described.	There is no plural for 'universe'.
The universe cannot be distinguished from vacuum.	The universe was not created.

ticular, a perfect physics book describes itself, its own production, its own author, its own readers and its own contents.

Since the universe is not a set, a perfect physics book *can* exist, as it does not contradict any property of the universe. But now a further question arises.

Does the universe make sense?

> “Drum hab ich mich der Magie ergeben,
> […]
> Daß ich erkenne, was die Welt
> Im Innersten zusammenhält.*”
>
> Goethe, *Faust*

Is the universe really the sum of matter–energy and space-time? Or of particles and vacuum? We have heard these statements so often that we may forget to check them. We do not need magic, as Faust thought: we only need to list what we have found so far, especially in this section, in the section on Planck scales, and in the chapter on brain and language. Table 3 shows the result. Vol. III, page 175

Not only are we unable to state that the universe is made of space-time and matter; we are unable to say anything about the universe at all!** It is not even possible to say that

---

[*] 'Thus I have devoted myself to magic, […] that I understand how the innermost world is held together.'

[**] There is another well-known, non-physical concept about which nothing can be said. Many scholars have explored it in detail. What is it? Challenge 76 s

it exists, since it is impossible to interact with it. The term 'universe' does not allow us to make a single sensible statement. (Can you find one?) Challenge 77 r We are only able to list properties it does *not* have. We are unable to find any property that the universe *does* have. Thus, the universe has no properties! We cannot even say whether the universe is something or nothing. *The universe isn't anything in particular.* In other words, the term 'universe' is not at all useful for the description of motion.

Vol. III, page 195 We can obtain a confirmation of this strange conclusion from an earlier chapter. There we found that any concept needs defined content, defined limits and a defined domain of application. In this section, we have found that the term 'universe' has none of these; there is thus no such concept. If somebody asks why the universe exists, the answer is: not only does the use of the word 'why' wrongly suggest that something may exist outside the universe, providing a reason for it and thus contradicting the definition of the term 'universe' itself; but more importantly, the universe does not exist, because there is no such concept as a 'universe'.

In summary, any sentence containing the word 'universe' is meaningless. The word only *seems* to express something, but it doesn't.* This conclusion may be interesting, even strangely beautiful, but does it help us to understand motion more precisely? Yes, it does.

Abandoning sets and discreteness eliminates contradictions

Our discussion of the term 'universe' shows that the term cannot include any element or set. Nature cannot be made of atoms. Nature cannot be made of space-time points. Nature cannot be made of separate, distinct and discrete entities.

The difficulties in giving a sharp definition of 'universe' also show that the fashionable term 'multiverse' makes no sense. There is no way to define such a term, since there is no empirical way and also no logical way to distinguish 'one' universe from 'another': the universe has no boundary.

By taking into account the limits on length, time, mass and all the other quantities we have encountered, we have reached a number of almost painful conclusions about nature. However, we have also received something in exchange: all the contradictions between general relativity and quantum theory that we mentioned at the beginning of this chapter are now resolved. Challenge 78 e *We changed the contradictions to circular definitions.* Although we have had to leave many cherished habits behind us, in exchange we have the promise of a description of nature without contradictions. But we get even more.

Extremal scales and open questions in physics

Page 18 At the beginning, we listed all the fundamental properties of nature that are unexplained either by general relativity or by quantum theory. We called it the millennium list. The results of this chapter provide us with surprising statements on many of the items. In fact, many of the statements are not new at all, but are surprisingly familiar. Let us com-

---

* Of course, the term 'universe' still makes sense if it is defined more restrictively: for example, as everything interacting with a particular human or animal observer in everyday life. But such a definition, equating 'universe' and 'environment', is not useful for our quest, as it lacks the precision required for a description of motion.

TABLE 4 Properties of nature at maximal, everyday and minimal scales

Physical property of nature	At horizon scale	At everyday scale	At Planck scale
requires quantum theory and relativity	true	false	true
intervals can be measured precisely	false	true	false
length and time intervals appear	limited	unlimited	limited
space-time is not continuous	true	false	true
points and events cannot be distinguished	true	false	true
space-time is not a manifold	true	false	true
space is 3-dimensional	false	true	false
space and time are indistinguishable	true	false	true
initial conditions make sense	false	true	false
space-time fluctuates	true	false	true
Lorentz and Poincaré symmetry	does not apply	applies	does not apply
CPT symmetry	does not apply	applies	does not apply
renormalization	does not apply	applies	does not apply
permutation symmetry	does not apply	applies	does not apply
interactions	do not exist	exist	do not exist
number of particles	undefined	defined	undefined
algebras of observables	undefined	defined	undefined
matter indistinguishable from vacuum	true	false	true
boundaries exist	false	true	false
nature is a set	false	true	false

pare systematically the statements from this chapter, on the universe, with those of the previous chapter, on Planck scales. The comparison is given in Table 4.

First, Table 4 shows that *each* unexplained property listed there is unexplained at *both* limits of nature, the small and the large limit. Worse, many of these unexplained general properties do not even *make sense* at the two limits of nature! However, there is hope.

Secondly, and more importantly, nature behaves in the *same way* at horizon scales and at Planck scales. In fact, we have not found any difference between the two cases. (Can you discover one?) We are thus led to the hypothesis that nature does not distinguish between the large and the small. Nature seems to be characterized by *extremal identity*.

Challenge 79 r

Is extremal identity a principle of nature?

The idea of extremal identity incorporates some rather general points:

- All open questions about nature appear at both size extremes.
- Any description of nature requires both general relativity and quantum theory.
- Nature, or the universe, is not a set.
- Initial conditions and evolution equations make no sense at nature's limits.

— There is a relation between local and global issues in nature.
— The concept of 'universe' has no content.

Extremal identity thus looks like a useful hypothesis in the search for a unified description of nature. To be a bit more provocative, it seems that extremal identity may be the *only* hypothesis incorporating the idea that the universe is not a set. Therefore, extremal identity seems to be essential in the quest for unification.

Challenge 80 e
Ref. 106

Extremal identity is beautiful in its simplicity, in its unexpectedness and in the richness of its consequences. You might enjoy exploring it by yourself. In fact, the exploration of extremal identity is currently the subject of much activity in theoretical physics, although often under different names.

The simplest approach to extremal identity – in fact, one that is too simple to be correct – is *inversion*. It looks as if extremal identity implies a connection such as

$$r \leftrightarrow \frac{l_{\mathrm{Pl}}^2}{r} \quad \text{or} \quad x_\mu \leftrightarrow \frac{l_{\mathrm{Pl}}^2\, x_\mu}{x_\mu\, x^\mu} \qquad (114)$$

Challenge 81 s

relating distances r or coordinates x_μ with their inverse values using the Planck length l_{Pl}. Can this mapping be a symmetry of nature? At every point of space? For example, if the horizon distance is inserted, the relation (114) implies that lengths smaller than $l_{\mathrm{Pl}}/10^{61} \approx 10^{-96}$ m never appear in physics. Is this the case? What would inversion imply for the big bang?

More involved approaches to extremal identity come under the name of *space-time duality* and *holography*. Numerous fascinating questions are contained in extremal identity; there is a lot of fun ahead of us.

Challenge 82 e

Above all, we need to find the correct version of the inversion relation (114). Inversion is neither sufficient nor correct. It is not sufficient because it does not explain *any* of the millennium issues left open by general relativity and quantum theory. It only *relates* some of them, but it does not *solve* any of them. (You may wish to check this for yourself.) In other words, we need to find the precise description of quantum geometry and of elementary particles.

Page 79

However, inversion is also simply wrong. Inversion is not the correct description of extremal identity because it does not realize a central result discovered above: it does not connect *states* and *intrinsic properties*, but keeps them distinct. In particular, inversion does not take *interactions* into account. And most open issues at this point of our mountain ascent concern the properties and the appearance of interactions.

SUMMARY ON THE UNIVERSE

In summary, we found some additional requirements for the final theory we are looking for. Whenever we combine general relativity and quantum theory, the universe teaches us that it is not a set of parts. For this reason, any sentence or expression containing the term 'universe' is probably meaningless, whenever complete precision is required.[*] We also learned that a description of nature without sets solves the contradictions between general relativity and quantum theory.

---

[*] For example, the term 'universe' cannot be the subject of a sentence, nor its object.

We then found an intriguing relation between Planck scales and cosmological scales: they seem to pose the same challenges to their description. There is a tight relation between large and small scales in nature. In short, there seems to be little difference – if any at all – between the universe and nothing.

A PHYSICAL APHORISM

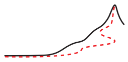
Ref. 54, Ref. 55

Despite these confusing results, we are approaching the top of Motion Mountain. To motivate us to continue, here is a humorous 'proof' that we near the top. Salecker and Wigner, and then Zimmerman, formulated the fundamental limit for the measurement precision τ attainable by a clock of mass M. It is given by $\tau = \sqrt{\hbar T / M c^2}$, where T is the time to be measured. We can then ask what time T can be measured with a precision of a Planck time t_{Pl}, given a clock of the mass of the whole universe. We get a maximum time of

$$T = \frac{t_{\mathrm{Pl}}^2 c^2}{\hbar} M \; . \qquad (115)$$

Challenge 83 e

Inserting numbers, we find rather precisely that the time T is the present age of the universe. With the right dose of humour we can see this result as a sign that time is now ripe, after so much waiting, for us to understand the universe down to the Planck scale. We are thus getting nearer to the top of Motion Mountain. Be prepared for a lot of fun. To find inspiration for the rest of our adventure, we take a break.



CHAPTER 6

THE PHYSICS OF LOVE – AN INTERMEDIATE REPORT

> “Sex is the physics urge sublimated.”
>
> Graffito

Maybe you have once met a physicist who has told you, in one of those moments of confidentiality, that studying physics is even more beautiful than making love. At this statement, many will simply shake their head in disbelief or disapproval. In this chapter we will argue that it is possible to learn so much about physics while making love that discussions about their relative beauty can be put aside altogether.

Imagine being with your partner on a beautiful tropical island, just after sunset, and looking together at the evening sky. Imagine, as well, that you know little of what is taught at school nowadays. To be specific, imagine that your knowledge is that of the late Renaissance (which is probably about the average level of modern scientific education anyway).

You are busy enjoying each other’s company. The most important results of physics can be deduced from the following experimental facts:*

Love is communication.
Love is an interaction.
Love is attractive.
Love makes noise.
Love is for reproduction.
Love needs memory.
Love uses the sense of sight.
Love is motion.
Love is based on touch.
Love is fun.

Love is tiring.
Love takes time.
Love is repulsive.
In love, size matters.
Love can hurt.
Love is Greek.
Love is animal.
Love is holy.
Love uses motion.
Love is private.
Love makes one dream.

Let us see what these observations imply for the description of nature.

Ref. 107

* Here we deduce physics from love. We could also deduce physics from sexuality. The original chapter title ‘The physics of sex’ was abandoned to stop shocking both the more sensible and the more indignant, i.e., the more violent characters.

* *

Love is *communication*. Communication is possible because nature is regular: it looks similar from different standpoints and because there are no surprises in nature. Without similarity we could not understand each other. A world of surprises would even make thinking impossible: it would not be possible to form concepts to describe observations. But fortunately, the world is regular; it thus allows us to use concepts such as time and space for its description.

* *

Love is an *interaction between moving bodies*. Together with the previous result, this implies that we can and need to describe moving bodies with *mass*, *energy* and *momentum*, and that motion can be predicted. That is not a small feat. For example, it implies that the Sun will rise tomorrow. Challenge 84 s

* *

Love is *attractive*. When feeling attracted to your partner, you may wonder if this attraction is the same as that which keeps the Moon going around the Earth. You make a quick calculation and find that applying the expression for universal gravity

$$E_{\text{pot}} = -\frac{GMm}{r} \qquad (116)$$

to both of you, the energy involved is about as much as the energy added by the leg of a fly on the skin. (M and m are your masses, r your distance apart, and the gravitational constant has the value $G = 6.7 \cdot 10^{-11}\,\mathrm{m^3/kg\,s^2}$.) In short, your partner teaches you that in nature there are other attractive interactions apart from gravity.

Nevertheless, this first equation is important: it allows us to predict the position of the planets, the time of the tides, the time of eclipses, the return of comets, and so on, to a high accuracy, thousands of years in advance.

* *

Love *makes noise*. That is not news. However, even after making love, even when everybody and everything is quiet, in a completely silent environment, we do hear something. The noises we hear are produced within the ear, partly by the blood flowing through the head, partly by the electrical noise generated in the nerves. That is strange. If matter were continuous, there would be no noise even for low signal levels. In fact, all proofs of the discreteness of matter, of electric current, of energy, or of light are based on the increase of fluctuations with the smallness of the system under consideration. The persistence of noise thus makes us suspect that matter is made of smallest entities. Making love confirms this suspicion in several ways.

* *

Love is for *reproduction*. We owe our life to love, as we are all results of reproduction. But the reproduction of a structure is possible only if it can be constructed, in other words if the structure can be built from small standard entities. Thus we again suspect ourselves to be made of smallest, discrete entities.

Love is also a complicated method of reproduction. Mathematics provides a much simpler one. If matter objects were not made of particles, but were continuous, it would be possible to perform reproduction by cutting and reassembling. A famous mathematical theorem by Banach and Tarski proves that it is possible to take a continuous solid, cut it into five pieces and rearrange the pieces to obtain *two* copies of the same size and volume as the original. In fact, volumes can be increased in this way, thus achieving growth without any need for food. Mathematics thus provides some interesting methods for growth and reproduction. However, these mathematical methods assume that matter is continuous, without a smallest length scale. The observation that these methods do not work in nature is compatible with the idea that matter is not continuous.

* *

Ref. 108 Love *needs memory*. If you could not recognize your partner among all possible ones, your love life would be quite complicated. A memory is a device which, in order to store information, must have *small* internal fluctuations. Fluctuations in systems get smaller as the number of components in them increases. Since our memory works so well, we must be made of a large number of small particles.

In short, love shows that we are made of some kind of Lego bricks: depending on the level of magnification, these bricks are called molecules, atoms, or elementary particles. It is possible to estimate their size using the sea around the tropical island and a bit of oil. Can you imagine how? Challenge 85 s

* *

Love *uses the sense of sight*. It is only possible for us to see each other because we are cold whereas the Sun is hot. If we and our environment all had the same temperature as the Sun, we would not see each other. This can be checked experimentally: looking into a glowing oven filled with glowing objects, it is impossible to discern the objects against the background. Ref. 109 Vol. III, page 120

* *

Love *is motion*. Lovers move. Moreover, their speed can be measured. Since measurement is a comparison with a standard, there must be a velocity standard in nature: some special velocity that stands out. Such a standard must be either the minimum or the maximum possible value. Now, we know from daily life that there is no minimum velocity, so we must look for a maximum value. To estimate the value of the maximum, just take out your mobile phone and ring home from the island to your family. From the delay in the line and the height of the satellite, you can deduce the telephone speed c and get $3 \cdot 10^8$ m/s.

The existence of a maximum speed c implies that time is different for different observers. Looking into the details, we find that this effect becomes noticeable at energies

$$E_{\text{different time}} \approx mc^2 , \tag{117}$$

where m is the mass of the object. For example, this applies to electrons inside a television tube.

* *

Love is based on *touch*. When we touch our partner, we sometimes get small shocks. The energies involved are larger than those of touching fly legs. In short, people are electric.

In the dark, we observe that discharges emit light. Light is thus related to electricity. In addition, touching proves that light is a wave: simply observe the dark lines between two fingers near your eye in front of a bright background. The lines are due to interference effects. Thus light does not move with infinite speed. In fact, it moves with the same speed as mobile phone calls. Challenge 86 s

* *

Love is *fun*. People like to make love in different ways – for example, in a dark room. But rooms get dark when the light is switched off only because we live in a space with an odd number of dimensions. In even-dimensional space, a lamp would not turn off directly after the switch is flicked, but only dim slowly. Vol. I, page 242

Love is also fun because with our legs, arms and bodies we can make knots. Knots are possible only in three dimensions. In short, love is fun only because we live in three dimensions.

* *

Love is *tiring*. The reason is gravity. But what is gravity? A little thinking shows that since there is a maximum speed, gravity is the curvature of space-time. Curved space also means that a *horizon* – a largest possible visible distance – can appear. From equations (116) and (117), we deduce that this happens when distances are of the order of

$$R_{\text{horizon}} \approx Gm/c^2 \ . \tag{118}$$

For example, it is only because of a horizon – albeit one appearing in a somewhat different way – that the night sky is dark.

* *

Love *takes time*. It is known that men and women have different opinions on durations. It is also known that love happens between your ears. Indeed, biological research has shown that we have a clock inside the brain, which depends on circulating electric currents. This clock provides our normal sense of time. Since this clock exists, there must be a time standard in nature. Such a standard must be either a minimum or a maximum time interval. We shall discover it later on. Ref. 110

* *

Love is *repulsive*. And in love, *size matters*. These facts turn out to be two sides of the same coin. Love is based on touch, and touch needs repulsion. Repulsion needs a length scale, but neither gravity nor classical electrodynamics provides one. Classical physics only allows for the measurement of speed. It cannot explain how the measurement of length, time, or mass is possible.* Classically, matter cannot be hard: it should be possible to compress it. But love shows us that this is not the case. Love shows us that lengths

---

* Even the classical electron radius depends on a length scale through the elementary charge e. Vol. II, page 254

scales do exist, and thus that classical physics is not sufficient for the description of nature.

* *

Love can *hurt*. For example, it can lead to injuries. Atoms can get ripped apart. That happens when energies are concentrated on small volumes: a few attojoule per atom. Investigating such situations more precisely, we finds that strange phenomena appear at distances r if energies exceed the value

$$E \approx \frac{\hbar\, c}{r} \quad . \tag{119}$$

In particular, energy becomes chunky, things become fuzzy, boxes are not tight, and particles get confused. These are called *quantum* phenomena. The new constant $\hbar = 10^{-34}$ Js is important: it determines the size of things, because it allows us to define distance and time units. In other words, objects tear and break because in nature there is a minimum action, given roughly by $\hbar$.

If even more energy is concentrated in small volumes – of the order of mc^2 per particle – one even observes transformation of energy into matter, or *pair production*. From equations (117) and (119), we deduce that this happens at distances of

$$r_{\text{pair production}} \approx \frac{\hbar}{m\, c} \; . \tag{120}$$

At such small distances we cannot avoid using the quantum description of nature.

* *

Love is not only *Greek*. The Greeks were the first to make theories above love, as Plato did in his *Phaedrus*. But they also described it in another way. Before Plato, Democritus said that making love is an example of particles moving and interacting in a vacuum. If we change 'a vacuum' to 'curved 3+1-dimensional space' and 'particle' to 'quantum particle', we do indeed make love in the way Democritus described 2,500 years ago.

It seems that physics has not made much progress in the mean time. Take the statement made in 1939 by the British astrophysicist Arthur Eddington:

Ref. 111

> I believe there are 15,747,724,136,275,002,577,605,653,961,181,555,468,044,717, 914,527,116,709,366,231,425,076,185,631,031,296 protons in the universe and the same number of electrons.

Compare it with the version of the year 2000:

> Baryons in the universe: $10^{80\pm1}$. Total charge: near zero.

The second version is more honest, but which of the two is more sensible? None is. Both statements show that the Greek description nature, leaves some observations unexplained, in particular the number of particles.

* *

Love is *animal*. We have seen that we can learn a lot about nature from the existence of love. We could be tempted to see this approach to understanding nature as a case of the so-called *anthropic principle*. However, some care is required here. We could have learned exactly the same if we had taken as our starting point the observation that *apes* or *pigs* have love. There is no 'law' of nature which distinguishes between them and humans. In fact, there is a simple way to determine whether any 'anthropic' statement makes sense: the reasoning must be equally true for humans, apes, and pigs.

One famous anthropic deduction was stated by the British astrophysicist Fred Hoyle. While studying stars, he predicted a resonance in the carbon-12 nucleus. If it did not exist, he argued, stars could not have produced the carbon which afterwards was spread out by explosions into interstellar space and collected on Earth. Since apes or pigs might also reason in this way, Hoyle's statement does make sense.

On the other hand, claiming that the universe is made *especially* for people is not sensible: pigs would say that it is made for pigs. The existence of either humans or pigs requires all 'laws' of nature. In summary, the anthropic principle is valid only in so far as its consequences are indistinguishable from those of the porcine or the simian principle. The animal side of love imposes strict limits on the philosophy of physics.

* *

Love is *holy*. According to the famous definition by the theologian Rudolf Otto, holiness results from a mixture of a *mysterium tremendum* and a *mysterium fascinans*. *Tremendum* means that it makes one tremble. Indeed, love produces heat and is a dissipative process. All systems in nature that produce heat have a finite lifetime. That is true for machines, stars, animals, lightning, fire, lamps and people. Through the involved heat, love shows us and remind us that we are going to die. Physicists call this the second principle of thermodynamics.

Ref. 112

But love also fascinates. Everything that fascinates has a story. Indeed, this is a principle of nature: every dissipative structure, every structure which appears or is sustained through the release of energy, has a story. Take atoms, for example. All the protons we are made of formed during the big bang. Most of the hydrogen we are made of is also that old. The other elements were formed in stars and then blown into the sky during nova or supernova explosions. They then regrouped during planet formation. We really are made of stardust.

Why do such stories fascinate? If you only think about how you and your partner met, you will discover that it was through a chain of incredible coincidences. If only one of these coincidences had not taken place, you and your partner would not be together. And of course, we all owe our existence to such chains of coincidences, which brought our parents together, and our grandparents, and made life appear on Earth.

The realization of the importance of coincidences automatically produces two kinds of questions: *Why* did they happen? And *what if* they had not happened? The game is especially interesting if after each answer, the questions are repeated. We all have discovered that there is an end to this repetition. Physicists have produced a list of answers to all the final 'why' questions, and many are working on the 'what if' questions. The first

Page 18 list, the 'why' list of Table 1, gives all those facts that were still unexplained in the year 2000. It could be called the *millennium list* of coincidences, or of unexplained facts, or of nature's challenges.

It is equally interesting to study what the consequences would be if any of the values from the 'why' list were just a tiny bit different. In turns out that even small changes in nature would lead to completely different observations. A short version of this 'what of' list is given in Table 5; a longer version is given in the volume on *Pleasure, Technology and Stars*. Vol. V, page 242

TABLE 5 A small selection of consequences of changing aspects of nature

Observable	Change	Result
Moon size	smaller	small Earth magnetic field; too much cosmic radiation; widespread child cancers
Moon size	larger	large Earth magnetic field; too little cosmic radiation; no evolution into humans
Jupiter	smaller	many comet impacts on Earth; extinction of animal life
Jupiter	larger	few comet impacts on Earth; no Moon; no dinosaur extinction
Oort belt	smaller	no comets, no irregular asteroids, no Moon; still dinosaurs
Star distance	smaller	irregular planetary motion; supernova dangers
Strong coupling constant	smaller	proton decay; leukaemia

The large number of coincidences of life forces our mind to realize that we are only a *tiny* part of nature. We are a small droplet shaken around in the ocean of nature. Even the tiniest changes in nature would prevent the existence of humans, apes and pigs. In other words, making love tells us that the universe is much larger than we are, and how much we are dependent on and connected to the rest of the universe.

* *

We said above that love uses *motion*. This fact contains a remarkable mystery, worth a second look.

Motion is the change of position with time of some bodies. Position is what we measure with a ruler. Time is what we measure with a clock. Both rulers and clocks are bodies.

A body is an entity distinct from its environment by its shape or its mass. Shape is the extension of a body in space (and time). Mass is measured by measuring speed or acceleration, i.e., by measuring space and time.

This means that we define space-time in terms of bodies – as in general relativity – and we define bodies in terms of space-time – as in quantum theory. This circular reasoning shows that making love is truly a mystery. But any attempt to eliminate the circular reasoning has to overcome major difficulties – which can also be experienced while making love.

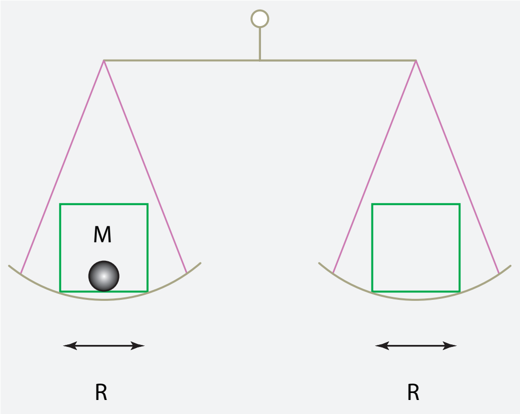


FIGURE 10 A Gedanken experiment showing that at Planck scales, matter and vacuum cannot be distinguished.

* *

Love is *private*. But is it? The concept of privacy assumes that a person can separate from the rest, without important interactions, at least for a given time, and come back later. This is possible only if there is enough *empty space* between the person and others. In other words, privacy is based on the idea that objects can be distinguished from vacuum. Let us check whether this is always possible.

Ref. 44 What is the smallest measurable distance? This question was *almost* answered by Max Planck in 1899. The distance δl between two objects of mass m is surely larger than their position indeterminacy $\hbar/\Delta p$; and the momentum indeterminacy must be smaller that the momentum leading to pair production, i.e., $\Delta p < mc$. Thus

$$\delta l \geqslant \Delta l \geqslant \frac{\hbar}{mc} . \qquad (121)$$

In addition, measurement requires that signals leave the objects, so the two masses must not be black holes. Their masses must be small enough to ensure that the Schwarzschild radius is smaller than the distance to be measured. This means that $r_S \approx Gm/c^2 < \delta l$, or

$$\delta l \geqslant \sqrt{\frac{\hbar G}{c^3}} = l_{\mathrm{Pl}} = 1.6 \cdot 10^{-35}\,\mathrm{m} . \qquad (122)$$

This expression defines a minimum length in nature, the so-called *Planck length*. Every other Gedanken experiment leads to this characteristic length as well. In fact, this minimum distance (and the corresponding minimum time interval) provides the measurement standard we were looking for at the beginning of our musings about length and time measurements.

In other words, privacy has its limits. In fact, the issue becomes even more muddled when we explore the consequences for bodies. A body (even a human one) is something we can touch, throw, hit, carry or weigh. Physicists say that a body is something with energy or momentum. Vacuum has none of these properties. In addition, vacuum is unbounded, whereas objects are bounded.

What happens if we try to weigh objects at Planck scales? Quantum theory makes a simple prediction. If we put an object of mass M in a box of size R onto a scale – as in Figure 10 – equation (119) implies that there is a minimal mass error ΔM given by

$$\Delta M \approx \frac{\hbar}{cR} \; . \tag{123}$$

If the box has Planck size, the mass error is the Planck mass

$$\Delta M = M_{\mathrm{Pl}} = \sqrt{\hbar c/G} \approx 22\,\mu\mathrm{g} \; . \tag{124}$$

How large a mass can we put into a box of Planck size? Obviously one with the maximum possible mass density. To determine this, imagine a planet with a satellite in orbit around it, just skimming its surface. The density ρ of the planet with radius r is given by

$$\rho \approx \frac{M}{r^3} = \frac{v^2}{Gr^2} \; . \tag{125}$$

Using equation (121) we find that the maximum mass density in nature, within a factor of order one, is the so-called *Planck density*, given by

$$\rho_{\mathrm{Pl}} = \frac{c^5}{G^2\hbar} = 5.2 \cdot 10^{96}\,\mathrm{kg/m}^3 \; . \tag{126}$$

Therefore the maximum mass that can be contained inside a Planck box is the Planck mass. But that was also the measurement *error* for that situation. This implies that we cannot say whether the original box we measured was empty or full: vacuum *cannot* be distinguished from matter at Planck scales. This astonishing result is confirmed by every other Gedanken experiment exploring this issue. Ref. 44

It is straightforward to deduce with similar arguments that objects are not bound in size at Planck scales, i.e., that they are not localized, and that the vacuum is not necessarily extended at those scales. In addition, the concept of particle number cannot be defined at Planck scales. Challenge 87 s

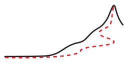
So, why is there something instead of nothing? Making love shows that there is no difference between the two options!

* *

Love makes us *dream*. When we dream, especially at night, we often look at the sky. How far away is it? How many atoms does it enclose? How old is it? These questions have an answer for small distances and for large distances; but for the whole of the sky, or the whole of nature, they cannot have an answer, as there is no way to be outside the sky in order to measure it. In fact, the impossibility of measuring nature at the smallest scales is found again at the largest scales. There seems to be a fundamental equivalence, or a *duality*, between the very large and the very small. Page 107

Summary on love and physics

In short, making love is a good physics lesson. Enjoy the rest of your day.



Chapter 7

THE SHAPE OF POINTS – EXTENSION IN NATURE

> Nil tam difficile est, quin quaerendo investigari possiet.
>
> Terence*

The usual expressions for the corrected Compton wavelength $\lambda = \hbar/mc$ and for the Schwarzschild radius $r_s = 2Gm/c^2$ imply a number of arguments which lead to the conclusion that at Planck energies, what we call 'space points' and 'point particles' must actually be described by *extended* constituents that are infinite and fluctuating in size. We will show this in the following ways.

1. Any experiment trying to measure the size or the shape of an elementary particle with high precision inevitably leads to the result that at least one dimension of the particle is of macroscopic size.
2. There is no evidence that empty space is continuous, but plenty of evidence that it is not. In particular, in order to build up an entity, such as the vacuum, that is extended in three dimensions, the constituents must be extended.
3. The existence of minimum measurable distances and time intervals implies the existence of space-time duality: a symmetry between very large and very small distances. Space-time duality in turn implies that the fundamental constituents that make up vacuum and matter are extended.
4. The constituents of the universe, and thus of vacuum, matter and radiation, cannot form a set. But any precise description of nature without sets must use extended constituents.
5. The Bekenstein–Hawking expression for the entropy of black holes – in particular its surface dependence – confirms that both vacuum and particles are composed of extended constituents.
6. The attempt to extend statistical properties to Planck scales shows that both particles and space points behave as braids at high energies, and thus as extended constituents.
7. The Dirac construction for spin provides a model for fermions that matches observations and points to extended constituents.

We finish with some experimental and theoretical checks of extension and an overview of present research efforts.

---

* 'Nothing is so difficult that it could not be investigated.' Terence is Publius Terentius Afer (*c.* 190–159 BCE), important roman poet. He writes this in his play *Heauton Timorumenos*, verse 675.

> “Also, die Aufgabe ist nicht zu sehen, was noch nie jemand gesehen hat, sondern über dasjenige was jeder schon gesehen hat zu denken was noch nie jemand gedacht hat.[*]”
>
> Erwin Schrödinger

THE SIZE AND SHAPE OF ELEMENTARY PARTICLES

Size is the length of vacuum taken by an object. This definition comes naturally in everyday life, quantum theory and relativity. To measure the size of an object as small as an elementary particle, we need high energy. The higher the energy, the higher the precision with which we can measure the size.

However, near the Planck energy, vacuum and matter cannot be distinguished: it is impossible to define the boundary between the two, and thus it is impossible to define the size of an object. As a consequence, every object, and in particular every elementary particle, becomes as extended as the vacuum! There is no measurement precision at all at Planck scales. Can we save the situation? Let us take a step back. Do measurements at least allow us to say whether particles can be contained inside small spheres?

Do boxes exist?

The first and simplest way to determine the size of a compact particle such as a sphere, or at least an upper limit, is to measure the size of a *box* it fits in. To be sure that the particle is inside, we must first be sure that the box is tight: that is, whether anything (such as matter or radiation) can leave the box.

But there is no way to ensure that a box has no holes! We know from quantum physics that any wall is a finite potential hill, and that tunnelling is always possible. In short, there is no way to make a completely tight box.

Let us cross-check this result. In everyday life, we call particles ‘small’ when they can be enclosed. Enclosure is possible in daily life because walls are impenetrable. But walls are only impenetrable for matter particles up to about 10 MeV and for photons up to about 10 keV. In fact, boxes do not even exist at medium energies. So we certainly cannot extend the idea of ‘box’ to Planck energy.

Since we cannot conclude that particles are of compact size by using boxes, we need to try other methods.

Can the Greeks help? – The limitations of knives

The Greeks deduced the existence of atoms by noting that matter cannot be divided indefinitely. Quantum physics confirmed the conclusion, but modified it: nowadays, the elementary particles are the ‘atoms’ of matter and radiation. However, at Planck energy, we have a different situation. Cutting is the insertion of a wall,. But a wall is never perfect, and thus knives are limited just as walls are. Any attempt to divide matter must stop when (or before) we reach Planck scales. At Planck energy, any subdivision is impossible.

---

[*] ‘Our task is not to see what nobody has ever seen, but to think what nobody has ever thought about that which everybody has seen already.’

The limitations of walls and knives imply that at Planck energy, a cut does not necessarily lead to two separate parts. We can never state that the two parts have been really, completely separated: the possibility of a thin connection between the two parts to the right and left of the blade can never be excluded. In short, we cannot prove compactness by cutting objects at Planck scales.

ARE CROSS SECTIONS FINITE?

To sum up: despite all attempts, we cannot show that elementary particles are point-like; in fact, elementary particles are not even compact. Are they, at least, of finite size?

To determine the size of a particle, we can try to determine its departure from point-likeness. Detecting this departure requires scattering. For example, we can suspend the particle in a trap and then shoot a probe at it. What happens in a scattering experiment at highest energies? This question has been studied by Leonard Susskind and his colleagues. When shooting at the particle with a high-energy probe, the scattering process is characterized by an interaction time. Extremely short interaction times imply sensitivity to the size and shape fluctuations due to the quantum of action. An extremely short interaction time also provides a cut-off for high-energy shape and size fluctuations, and thus determines the measured size. As a result, the size measured for any microscopic, but extended, object *increases* when the probe energy is increased towards the Planck value.

Ref. 114

In summary, even though at experimentally achievable energies the size of an elementary particle is always smaller than the measurement limit, when we approach the Planck energy, the particle size increases above all bounds. So at high energies we cannot give an upper limit to the size of a particle. In other words, since particles are not point-like at everyday energies, at Planck energy they are enormous: particles are extended.

Particles are extended. That is quite a statement. Right at the start of our mountain ascent, we distinguished objects from their environment. Objects are by definition localized, bounded and compact. All objects have a *boundary*, i.e., a surface which does not itself have a boundary. Objects are also bounded in abstract ways: also the set of symmetries of an object, such as its geometric symmetry group or its gauge group, is bounded. In contrast, the environment is not localized, but extended and unbounded.

Vol. I, page 24

These basic assumptions fail us at Planck scales. At Planck energy, it is impossible to determine whether something is bounded or compact. Compactness and localization are only approximate properties; they are not applicable at high energies. In particular, the idea of a point particle is an approximate concept, valid only at low energies.

Let us perform another check on our conclusion that particles at Planck scales are as extended as the vacuum.

CAN WE TAKE A PHOTOGRAPH OF A POINT?

> “Καιρὸν γνῶθι.*”
>
> Pittacus

---

* ‘Recognize the right moment.’ Also rendered as: ‘Recognize thine opportunity.’ Pittacus (Πιττακος) of Mytilene (*c.* 650–570 BCE), was a Lesbian tyrant and lawmaker; he was also one of the ‘Seven Sages’ of ancient Greece.

TABLE 6 Effects of various camera shutter times on photographs

Duration	Blur	Observation possibilities and effects
1 h	high	Ability to see faint quasars at night if motion is compensated
1 s	high	Everyday motion completely blurred
20 ms	lower	Interruption by eyelids; small changes impossible to see
10 ms	lower	Effective eye/brain shutter time; tennis ball impossible to see while hitting it
0.25 ms	lower	Shortest commercial photographic camera shutter time; ability to photograph fast cars
1 µs	very low	Ability to photograph flying bullets; strong flashlight required
c. 10 ps	lowest	Study of molecular processes; ability to photograph flying light pulses; laser light required to get sufficient illumination
10 fs	higher	Light photography impossible because of wave effects
100 zs	high	X-ray photography impossible; only γ-ray imaging left over
shorter times	very high	Photographs get darker as illumination decreases; gravitational effects significant
10^{-43} s	highest	Imaging impossible

Humans – or any other types of observers – can only observe the world with finite resolution in time and in space. In this respect, humans resemble a film camera. Every camera has a resolution limit: it can only distinguish two events if they are a certain minimum distance apart and separated by a certain minimum time. What is the best resolution possible? The value was (almost) discovered in 1899: the Planck time and the Planck length. No human, no film camera and no apparatus can measure space or time intervals smaller than the Planck values. But what would happen if we took photographs with shutter times that *approach* the Planck time? Ref. 48 Ref. 44, Ref. 20

Imagine that you have the world's best shutter and that you are taking photographs at shorter and shorter times. Table 6 gives a rough overview of the possibilities. At shorter shutter times, photographs get darker. When the shutter time reaches the oscillation time of light, strange things happen: light has no chance to pass undisturbed; signal and noise become indistinguishable; and the moving shutter will produce colour shifts. In contrast to our everyday experience, the photograph would get more *blurred* at extremely short shutter times. Photography is impossible not only at long but also at short shutter times.

The difficulty of taking photographs is independent of the wavelength used. The limits move, but do not disappear. With a shutter time of τ, photons of energy lower than $\hbar/\tau$ cannot pass the shutter undisturbed. The blur is small when shutter times are those of everyday life, but *increases* when shutter times are shortened towards Planck times. As a result, there is no way to detect or confirm the existence of point objects by taking pictures. Points in space, as well as instants of time, are *imagined* concepts: they do not belong in a precise description of nature.

At Planck shutter times, only signals with Planck energy can pass through the shutter. Since at these energies matter cannot be distinguished from radiation or from empty space, all objects, light, and vacuum look the same. It is impossible to say what nature looks like at very short times.

But the situation is worse than this: a Planck shutter cannot exist at all, as it would need to be as small as a Planck length. A camera using it could not be built, as lenses do not work at this energy. Not even a camera obscura – without any lens – would work, as diffraction effects would make image production impossible.

In other words, the idea that at short shutter times a photograph of nature shows a frozen image of everyday life, like a stopped film, is completely wrong.* Indeed, at a single instant of time nature is not frozen at all. Zeno criticized this idea in his discussions of motion, though not as clearly as we can do now. At short times, nature is blurred. In particular, point particles do not exist.

In summary, whatever the intrinsic shape of what we call a 'point' might be, we know that, being always blurred, it is first of all a cloud. Whatever method is used to photograph an elementary particle, the picture is always extended. Therefore we study the shape in more detail.

WHAT IS THE SHAPE OF AN ELECTRON?

Since particles are not point-like, they have a shape. How can we determine it? We determine the shape of an everyday object by *touching* it from all sides. This works with plants, people or machines. It even works with molecules, such as water molecules. We can put them (almost) at rest, for example in ice, and then scatter small particles off them. Scattering is just a higher-energy version of touching. However, scattering cannot determine shapes of objects smaller than the wavelength of the probes used. To determine the shape of an object as small as an electron, we need the highest energies available. But we already know what happens when approaching Planck scales: the shape of a particle becomes the shape of all the space surrounding it. In short, the shape of an electron cannot be determined in this way.

Another way to determine the shape is to build a tight box around the system under investigation and fill it with molten wax. We then let the wax cool and observe the hollow part. However, near Planck energy, boxes do not exist. We are unable to determine the shape in this way.

A third way to measure the shape of an object is to cut it into pieces and then study the pieces. As is well known, the term 'atom' just means 'uncuttable' or 'indivisible'. However, neither atoms nor indivisible particles can exist. Indeed, *cutting* is just a low-energy version of a scattering process. And the process does not work at high energies. Therefore, there is no way to prove that an object is indivisible at Planck scales. Our everyday intuition leads us completely astray at Planck energy.

We could try to distinguish transverse and longitudinal shape, with respect to the direction of motion. However, for transverse shape we get the same issues as for scattering; transverse shape diverges for high energy. And to determine longitudinal shape, we need at least two infinitely high potential walls. We already know that this is impossible.

A further, indirect way of measuring shapes is to measure the moment of inertia. A finite moment of inertia means a compact, finite shape. But when the measurement energy is increased towards Planck scales, rotation, linear motion and exchange become mixed up. We do not get meaningful results. Ref. 44

---

* In fact, a shutter does not exist even at medium energy: shutters, like walls, stop existing at around 10 MeV.

Yet another way to determine shapes is to measure the *entropy* of a collection of particles we want to study. This allows us to determine the dimensionality and the number of internal degrees of freedom. But at high energies, a collection of electrons would become a black hole. We will study this issue separately below, but again we find no new information.

Are these arguments watertight? We assumed three dimensions at all scales, and that the shape of the particle itself is fixed. Maybe these assumptions are not valid at Planck scales? Let us check the alternatives. We have already shown that because of the fundamental measurement limits, the dimensionality of space-time cannot be determined at Planck scales. Even if we could build perfect three-dimensional boxes, holes could remain in other dimensions. It does not take long to see that all the arguments against compactness work even if space-time has additional dimensions.

Ref. 44

IS THE SHAPE OF AN ELECTRON FIXED?

Only an object composed of localized constituents, such as a house or a molecule, can have a fixed shape. The smaller the system, the more quantum fluctuations play a role. No small entity of finite size – in particular, no elementary particle – can have a fixed shape. In every Gedanken experiment involving a finite shape, the shape itself fluctuates. But we can say more.

The distinction between particles and environment rests on the idea that particles have *intrinsic* properties. In fact, all intrinsic properties, such as spin, mass, charge, and parity, are localized. But we have seen that no intrinsic property is measurable or definable at Planck scales. Thus it is impossible to distinguish particles from the environment. In addition, at Planck energy particles have all the properties that the environment has. In particular, particles are extended.

In short, we cannot prove by experiments that at Planck energy elementary particles are finite in size in all directions. In fact, all experiments we can think of are compatible with extended particles, with ‘infinite’ size. More precisely, a particle always reaches the borders of the region of space-time under exploration. In simple words, we can also say that particles have *tails*.

Not only are particles extended, but their shape cannot be determined by the methods just explored. The only remaining possibility is that suggested by quantum theory: *the shape of a particle fluctuates.*

We reach the same conclusions for radiation particles. The box argument shows that radiation particles are also extended and fluctuating.

Incidentally, we have also settled an important question about *elementary* particles. We have already seen that any particle that is smaller than its own Compton wavelength must be elementary. If it were composite, there would be a lighter component inside it; this lighter particle would have a larger Compton wavelength than the composite particle. This is impossible, since the size of a composite particle must be larger than the Compton wavelength of its components.*

Vol. IV, page 91

---

* Examples are the neutron, positronium, or the atoms. Note that the argument does not change when the elementary particle itself is unstable, like the muon. The possibility that all components are heavier than the composite, which would avoid this argument, does not seem to lead to satisfying physical properties: for example, it leads to intrinsically unstable composites.

However, an elementary particle *can* have constituents, provided that they are not compact. The difficulties of compact constituents were described by Sakharov in the 1960s. If the constituents are extended, the previous argument does not apply, as extended constituents have no localized mass. As a result, if a flying arrow – Zeno's famous example – is made of extended constituents, it cannot be said to be at a given position at a given time. Shortening the observation time towards the Planck time makes an arrow disappear in the cloud that makes up space-time.* Ref. 24

SUMMARY OF THE FIRST ARGUMENT FOR EXTENSION

In summary, point particles do not exist at Planck scales. At Planck scales, all thought experiments suggest that matter and radiation are made of *extended and fluctuating constituents of infinite size*.

We note directly that for extended constituents the requirement of a non-local description is satisfied; in addition, for fluctuating constituents the requirement of a statistical description of the vacuum is satisfied. The argument forbidding composition of elementary particles is circumvented, as extended constituents have no mass. Thus the concept of Compton wavelength cannot be defined or applied to extended constituents, and elementary particles can have constituents if the constituents are extended. But if the constituents are extended, how can compact, point-like particles be formed from them? We will look at a few options shortly. Ref. 44 Challenge 88 e

THE SHAPE OF POINTS IN VACUUM

> Thus, since there is an impossibility that [finite] quantities are built from contacts and points, it is necessary that there be indivisible material elements and [finite] quantities.
> Aristotle, *Of Generation and Corruption*

Ref. 116

We are used to the idea that empty space is made of spatial points. However, at Planck scales, no measurement can give zero length, zero mass, zero area or zero volume. There is no way to state that something in nature is a point without contradicting experimental results.

Furthermore, the idea of a point is an extrapolation of what is found in small empty boxes getting smaller and smaller. But we have just seen that at high energies small boxes cannot be said to be empty. In fact, boxes do not exist at all, as they can never have impenetrable walls at high energies.

Also, the idea of a point as a limiting subdivision of empty space is untenable. At small distances, space cannot be subdivided, as division requires some sort of dividing wall, which is impossible.

Even the idea of repeatedly putting a point between two others cannot be applied. At high energy, it is impossible to say whether a point is exactly on the line connecting the outer two points; and near Planck energy, there is no way to find a point between them at all. In fact, the term 'in between' makes no sense at Planck scales.

---

* Thus at Planck scales there is no quantum Zeno effect. Ref. 115

We thus find that space points do not exist, just as point particles do not exist. But there are other reasons why space cannot be made of points. In order to form space. points need to be kept *apart* somehow. Indeed, mathematicians have a strong argument for why physical space cannot be made of mathematical points: the properties of mathematical spaces described by the Banach–Tarski paradox are quite different from those of the physical vacuum. The Banach–Tarski paradox states that a sphere made of mathematical points can be cut into five pieces which can be reassembled into two spheres each of the same volume as the original sphere. Mathematically, there are sets of points for which the concept of volume makes no sense. Physically speaking, we conclude that the concept of volume does not exist for continuous space; it is only definable if an *intrinsic length* exists. This is the case for matter and for vacuum. But any concept with an intrinsic length must be described by one or several extended constituents.* In summary, in order to build up space, we need *extended* constituents.

Ref. 44

Also the number of space dimensions is problematic. Mathematically, it is impossible to define the dimension of a set of points on the basis of the set structure alone. Any compact one-dimensional set has as many points as any compact three-dimensional set – indeed, as any compact set of any dimensionality greater than zero. To build up the *physical* three-dimensional vacuum, we need constituents that organize their neighbourhood. The fundamental constituents must possess some sort of ability to form bonds, which will construct or fill precisely three dimensions. Bonds require extended constituents. A collection of tangled constituents extending to the maximum scale of the region under consideration would work perfectly. Of course, the precise shape of the fundamental constituents is not yet known. In any case, we again find that any constituents of physical three-dimensional space must be *extended*.

In summary, we need extension to define dimensionality and to define volume. This is not surprising. We deduced above that the constituents of particles are extended. Since vacuum is not distinguishable from matter, we would expect the constituents of vacuum to be extended as well. Stated simply, if elementary particles are not point-like, then points in the vacuum cannot be either.

MEASURING THE VOID

To check whether the constituents of the vacuum are extended, let us perform a few additional Gedanken experiments. First, let us measure the size of a point in space. The clearest definition of size is in terms of the cross section. How can we determine the cross section of a point? We can determine the cross section of a piece of vacuum and then determine the number of points inside it. However, at Planck energy, we get a simple result: the cross section of a volume of empty space is independent of depth. At Planck energy, vacuum has a surface, but no depth. In other words, at Planck energy we can only state that a Planck layer covers the surface of a region. We cannot say anything about its interior. One way to picture this result is to say that what we call 'space points' are in fact long tubes.

---

* Imagining the vacuum as a collection of compact constituents, such as spheres, with Planck size in all directions would avoid the Banach–Tarski paradox, but would not allow us to deduce the number of dimensions of space and time. It would also contradict all the other results of this section. Therefore we do not explore it further.

Ref. 117
Challenge 89 s

Another way to determine the size of a point is to count the points found in a given volume of space-time. One approach is to count the possible positions of a point particle in a volume. However, at Planck energy point particles are extended and indistinguishable from vacuum. At Planck energy, the number of points is given by the surface area of the volume divided by the Planck area. Again, the surface dependence suggests that particles are long tubes.

WHAT IS THE MAXIMUM NUMBER OF PARTICLES THAT FIT INSIDE A PIECE OF VACUUM?

Another approach to counting the number of points in a volume is to fill a piece of vacuum with point particles.

The maximum mass that fits into a piece of vacuum is a black hole. But in this case too, the maximum mass depends only on the *surface* of the given region of vacuum. The maximum mass increases less rapidly than the volume. In other words, the number of physical points inside a region of space is only proportional to the surface area of the region. We are forced to conclude that vacuum must be made of extended constituents crossing the whole region, independently of its shape.

SUMMARY OF THE SECOND ARGUMENT FOR EXTENSION

In summary, Planck scales imply that *space is made of extended constituents of infinite size.* Space is not made of points, but of a web.

Vol. I, page 270
Ref. 118

More than two thousand years ago, the Greeks argued that matter must be made of particles because salt can be dissolved in water and because fish can swim through water. Now that we know more about Planck scales, we have to reconsider this argument. Like fish swimming through water, particles can move through vacuum; but since vacuum has no bounds and cannot be distinguished from matter, vacuum cannot be made of localised particles. However, another possibility allows for motion of particles through a vacuum: both vacuum and particles might be made of a web of extended constituents. Let us study this possibility in more detail.

THE LARGE, THE SMALL AND THEIR CONNECTION

> “I could be bounded in a nutshell and count myself a king of infinite space, were it not that I have bad dreams.”
>
> William Shakespeare, *Hamlet*

If two observables cannot be distinguished, there is a symmetry transformation connecting them. For example, by a change of observation frame, an electric field may change into a magnetic one. A symmetry transformation means that we can change the viewpoint (i.e., the frame of observation) in such a way that the same observation is described by one quantity from one viewpoint and by the corresponding quantity from the other viewpoint.

When measuring a length at Planck scales it is impossible to say whether we are measuring the length of a piece of vacuum, the Compton wavelength of a body, or the

Schwarzschild diameter of a body. For example, the maximum size for an elementary object is its Compton wavelength. The minimum size for an elementary object is its Schwarzschild radius. The actual size of an elementary object is somewhere in between. If we want to measure the size precisely, we have to go to Planck energy; but then all these quantities are the same. In other words, at Planck scales, there is a symmetry transformation between Compton wavelength and Schwarzschild radius. In short, *at Planck scales there is a symmetry between mass and inverse mass.*

As a further consequence, at Planck scales there is a symmetry between size and inverse size. Matter–vacuum indistinguishability means that there is a symmetry between length and inverse length at Planck energy. This symmetry is called *space-time duality* or *T-duality* in the literature of superstrings.* Space-time duality is a symmetry between situations at scale $n\, l_{\text{Pl}}$ and at scale $f\, l_{\text{Pl}}/n$, or, in other words, between R and $(f\, l_{\text{Pl}})^2/R$, where the number f is conjectured to have a value somewhere between 1 and 1000. Ref. 119

Duality is a genuine non-perturbative effect. It does not exist at low energy, since duality automatically also relates energies E and $E_{\text{Pl}}^2/E = \hbar c^3/GE$, i.e., it relates energies below and above Planck scale. Duality is not evident in everyday life. It is a quantum symmetry, as it includes Planck's constant in its definition. It is also a general-relativistic effect, as it includes the gravitational constant and the speed of light. Let us study duality in more detail.

Is small large?

> [Zeno of Elea maintained:] If the existing are many, it is necessary that they are at the same time small and large, so small to have no size, and so large to be without limits.
>
> Simplicius

Ref. 120

To explore the consequences of duality, we can compare it to rotational symmetry in everyday life. Every object in daily life is symmetrical under a full rotation of 2π. For the rotation of an observer, angles make sense only as long as they are smaller than 2π. If a rotating observer were to insist on distinguishing angles of $0, 2\pi, 4\pi$ etc., he would get a new copy of the universe at each full turn.

Similarly, in nature, scales R and l_{Pl}^2/R cannot be distinguished. Lengths make no sense when they are smaller than l_{Pl}. If, however, we insist on using even smaller values and on distinguishing them from large ones, we get a new copy of the universe at those small scales. Such an insistence is part of the standard continuum description of motion, where it is assumed that space and time are described by the real numbers, which are defined over arbitrarily small intervals. Whenever the (approximate) continuum description with infinite extension is used, the $R \leftrightarrow l_{\text{Pl}}^2/R$ symmetry pops up.

Duality implies that diffeomorphism invariance is only valid at medium scales, not at extremal ones. At extremal scales, quantum theory has to be taken into account in the proper manner. We do not yet know how to do this.

Space-time duality means that introducing lengths smaller than the Planck length (as when one defines space points, which have size zero) means at the same time introducing

---

* There is also an *S-duality*, which connects large and small coupling constants, and a *U-duality*, which is the combination of S- and T-duality.

things with very large ('infinite') value. Space-time duality means that for every small enough sphere the inside equals the outside.

Duality means that if a system has a small dimension, it also has a large one, and vice versa. There are thus no small objects in nature. So space-time duality is consistent with the idea that the basic constituents are extended.

UNIFICATION AND TOTAL SYMMETRY

Above, we have shown that at Planck energy, time and length cannot be distinguished, and that vacuum and matter cannot be distinguished. Duality shows that mass and inverse mass cannot be distinguished. As a consequence, we deduce that length, time, and mass cannot be distinguished from each other at *all* energies and scales! And since every observable is a combination of length, mass and time, *space-time duality means that there is a symmetry between all observables*. We call it the *total symmetry*.[*]

Total symmetry implies that there are many specific types of duality, one for each pair of quantities under investigation. Indeed, the number of duality types discovered is increasing every year. It includes, among others, the famous electric–magnetic duality we first encountered in electrodynamics, coupling constant duality, surface–volume duality, space-time duality, and many more. All this confirms that there is an enormous amount of symmetry at Planck scales. In fact, similar symmetries have been known right from the beginning of research in quantum gravity.

Vol. III, page 77

Ref. 121

Ref. 119

Most importantly, total symmetry implies that gravity can be seen as equivalent to all other forces. Space-time duality thus shows that unification is possible. Physicists have always dreamt about unification. Duality tells us that this dream can indeed be realized.

It may seem that total symmetry completely contradicts what was said in the previous section, where we argued that all symmetries are lost at Planck scales. Which result is correct? Obviously, both of them are.

At Planck scales, all low-energy symmetries are indeed lost. In fact, all symmetries that imply a *fixed* energy are lost. However, duality and its generalizations combine both small and large dimensions, or large and small energies. Most of the standard symmetries of physics, such as gauge, permutation and space-time symmetries, are valid at each fixed energy separately. But nature is not made this way. The precise description of nature requires us to take into consideration large and small energies at the same time. In everyday life, we do not do that. The physics of everyday life is an approximation to nature valid at low and fixed energies. For most of the twentieth century, physicists tried to reach higher and higher energies. We believed that precision increases with increasing energy. But when we combine quantum theory and gravity we are forced to change this approach. To achieve high precision, we must take high and low energy into account at the same time.[**]

---

[*] A symmetry between size and Schwarzschild radius, i.e., a symmetry between length and mass, will lead to general relativity. Additionally, at Planck energy there is a symmetry between size and Compton wavelength. In other words, there is a symmetry between length and inverse mass. This means that there is a symmetry between coordinates and wave functions. This is a symmetry between states and observables. It leads to quantum theory.

[**] Renormalization energy does connect different energies, but not in the correct way; in particular, it does not include duality.

The great differences between the phenomena that occur at low and high energies are the main reason why unification is so difficult. We are used to dividing nature along a scale of energies: high-energy physics, atomic physics, chemistry, biology, and so on. But we are not allowed to think in this way any more. We have to take all energies into account at the same time. That is not easy, but we do not have to despair. Important conceptual progress was made in the last decade of the twentieth century. In particular, we now know that we need only *one constituent* for all things that can be measured.

Since there is only one constituent, total symmetry is automatically satisfied. And since there is only one constituent, there are many ways to study it. We can start from any (low-energy) concept in physics and explore how it looks and behaves when we approach Planck scales. In the present section, we are looking at the concept of 'point'. Obviously, the conclusions must be the same whatever concept we start with, be it electric field, spin, or any other. Such studies thus provide a check for the results in this section. Challenge 90 d

SUMMARY OF THE THIRD ARGUMENT FOR EXTENSION

Unification implies thinking in terms of duality and the concepts that follow from it. The large and the small are connected. Duality points to one single type of extended constituents that defines all physical observables.

We still need to understand exactly what happens to duality when we restrict ourselves to low energies, as we do in everyday life. We explore this now. Challenge 91 e

DOES NATURE HAVE PARTS?

> “Pluralitas non est ponenda sine necessitate.*”
> William of Occam

Another argument, independent of those given so far, points towards a model of nature based on extended constituents. We know that any concept for which we can distinguish parts is described by a set. We usually describe nature as a set of objects, positions, instants and so on. The most famous set-theoretic description of nature is the oldest known, given by Democritus: Ref. 122

> The world is made of indivisible particles and void.

This description was extremely successful in the past: there are no discrepancies with observations. However, after 2500 years, the conceptual difficulties of this approach are obvious.

We know that Democritus was wrong, first of all, because vacuum and matter cannot be distinguished at Planck scales. Thus the word 'and' in his sentence is already a mistake.

---

* 'Multitude should not be introduced without necessity.' This famous principle is commonly called *Occam's razor*. William of Ockham (b. 1285/1295 Ockham, d. 1349/50 München), or Occam in the common Latin spelling, was one of the great thinkers of his time.In his famous statement he expresses that only those concepts which are strictly necessary should be introduced to explain observations. It can be seen as the requirement to abandon *beliefs* when talking about nature. But at this stage of our mountain ascent it has an even more direct interpretation.

Secondly, because of the existence of minimal scales, the void cannot be made of 'points', as we usually assume. Thirdly, the description fails because particles are not compact objects. Finally, total symmetry implies that we cannot distinguish parts in nature. Nothing can be distinguished from anything else with complete precision, and thus the particles or points in space that make up the naive model of the world cannot exist.

In summary, quantum theory and general relativity together show that in nature, *all partitions and all differences are only approximate.* Nothing can really be distinguished from anything else with complete precision. In other words, there is no way to define a 'part' of nature, whether for matter, space, time, or radiation. *Nature cannot be a set.*

The conclusion that nature is not a set does not come as a surprise. We have already encountered another reason to doubt that nature is a set. Whatever definition we use for the term 'particle', Democritus cannot be correct for a purely logical reason. The description he provided is *not complete.* Every description of nature that defines nature as a set of parts misses certain aspects. Most importantly, it misses the *number* of these parts. In particular, the number of particles and the number of dimensions of space-time must be specified if we describe nature as made from particles and vacuum. For example, we saw that it is rather dangerous to make fun of the famous statement by Arthur Eddington

Vol. III, page 236

Ref. 111

> I believe there are 15,747,724,136,275,002,577,605,653,961,181,555,468,044, 717,914,527,116,709,366,231,425,076,185,631,031,296 protons in the universe and the same number of electrons.

In fact, practically all physicists share this belief, although they usually either pretend to favour some other number, or worse, keep the number unspecified. We have seen during our walk that in modern physics many specialized sets are used to describe nature. We have used vector spaces, linear spaces, topological spaces and Hilbert spaces. But we consistently refrained, like all physicists, from asking about the origin of their sizes (mathematically speaking, of their dimensionality or cardinality). In fact, it is just as unsatisfying to say that the universe contains some specific number of atoms as it is to say that space-time is made of point-like events arranged in 3 + 1 dimensions. Both are statements about set sizes, in the widest sense. In a complete, unified description of nature the number of smallest particles and the number of space-time points must not be fixed beforehand, but must *result* from the description.

Any part of nature is by definition smaller than the whole of nature, and different from other parts. As a result, no description of nature by a set can possibly yield the number of particles or the dimensionality of space-time. As long as we insist on using space-time or Hilbert spaces for the description of nature, we *cannot* understand the number of dimensions or the number of particles.

That is not too bad, as we know already that nature is *not* made of parts. We know that parts are only approximate concepts. In short, if nature were made of parts, it could not be a unity, or a 'one.' On the other hand, if nature is a unity, it cannot have parts.* Nature cannot be separable exactly. It cannot be made of particles.

Ref. 123

* As a curiosity, practically the same discussion can already be found in Plato's *Parmenides*, written in the fourth century BCE. There, Plato musically ponders different arguments on whether nature is or can be a *unity* or a *multiplicity*, i.e., a set. It seems that the text is based on the real visit to Athens by Parmenides

To sum up, nature cannot be a set. Sets are lists of distinguishable elements. When general relativity and quantum theory are unified, nature shows no elements: *nature stops being a set at Planck scales*. This result clarifies a discussion we started earlier in relation to classical physics. There we discovered that matter objects were defined using space and time, and that space and time were defined using objects. Along with the results of quantum theory, this implies that in modern physics particles are defined in terms of the vacuum and the vacuum in terms of particles. This is clearly not a good idea. But we have just seen that since the two concepts are indistinguishable from each other, we cannot define them in terms of each other. Everything is the same. In fact, there is no 'every' and no 'thing'. Since nature is not a set, the circular reasoning is dissolved.

Space-time duality also implies that space is not a set. It implies that events cannot be distinguished from each other, and thus do not form elements of some space. Phil Gibbs has given the name *event symmetry* to this property of nature. This thought-provoking term, although still containing the term 'event', emphasizes the fact that it is impossible to use a set to describe space-time.

Ref. 124

In short, nature cannot be made of vacuum and particles. This is a bizarre result. Atomists, from Democritus to Galileo, have been persecuted throughout history. Were their battles all in vain? Let us continue to clarify our thoughts.

Does the universe contain anything?

To state that the universe contains something implies that we are able to distinguish the universe from its contents. However, we now know that precise distinctions are impossible. If nature is not made of parts, it is wrong to say that the universe *contains* something.

Let us go further. We need a description of nature that allows us to state that at Planck energy nothing can be distinguished from anything else. For example, it must be impossible to distinguish particles from each other or from the vacuum. There is only one solution: everything – or at least, what we call 'everything' in everyday life – must be made of the same single constituent. All particles are made of one 'piece'. Every point in space, every event, every particle and every instant of time must be made of the same single constituent.

An amoeba

> “A theory of everything describing nothing is not better than a theory of nothing describing everything.”
>
> Anonymous

We have found that parts are approximate concepts. The parts of nature are not strictly smaller than nature itself. As a result, any 'part' must be extended. Let us try to extract some more information about the constituents of nature.

In any unified theory, all the concepts that appear must be only *approximately* parts of the whole. Thus we need an entity Ω, describing nature, which is not a set but which can be approximated by one. This is strange. We are all convinced very early in our lives

---

and Zeno. (Their home city, Elea, was near Naples.) Plato does not reach a conclusion. Modern physics, however, does.

that we are a *part* of nature. Our senses provide us with this information. We are not used to thinking otherwise. But now we have to.

Let us straight away eliminate a few options for Ω. One concept without parts is the empty set. Perhaps we need to construct a description of nature from the empty set? We could be inspired by the usual construction of the natural numbers from the empty set. However, the empty set makes only sense as the opposite of some full set. So the empty set is not a candidate for Ω.

Vol. III, page 202

Another possible way to define approximate parts is to construct them from multiple copies of Ω. But in this way we would introduce a new set through the back door. Furthermore, new concepts defined in this way would not be approximate.

We need to be more imaginative. How can we describe a whole which has no parts, but which has parts approximately? Let us recapitulate. The world must be described by a single entity, sharing all properties of the world, but which can be approximated as a set of parts. For example, the approximation should yield a set of space points and a set of particles. But also, whenever we look at any 'part' of nature, without any approximation, we should not be able to distinguish it from the whole world. Composite objects are not always larger than their constituents. On the other hand, composed objects must usually *appear* to be larger than their constituents. For example, space 'points' or 'point' particles are tiny, even though they are only approximations. Which concept without boundaries can be at their origin? Using usual concepts, the world is everywhere at the same time; if nature is to be described by a single constituent, this entity must be extended.

The entity has to be a single one, but it must *seem* to be multiple: it has to be multiple approximately, as nature shows multiple aspects. The entity must be something folded. It must be possible to count the folds, but only approximately. (An analogy is the question of how many grooves there are on an LP or a CD: depending on the point of view, local or global, one gets different answers.) Counting folds would correspond to a length measurement.

The simplest model would be a single entity which is extended and fluctuating, goes to infinity, and allows approximate localization, thus allowing approximate definition of parts and points.* In more vivid imagery, nature could be described by some deformable, folded and tangled entity: a giant, knotted amoeba. An amoeba slides between the fingers whenever one tries to grab a part of it. A perfect amoeba flows around any knife trying to cut it. The only way to hold it would be to grab it in its entirety. However, for someone himself made of amoeba strands, this is impossible. He can only grab it approximately, by catching part of it and approximately blocking it, for example using a small hole, so that the escape takes a long time.

Summary of the fourth argument for extension

The lack of particles and of sets in nature leads to describing nature by a single constituent. Nature is thus modelled by an entity which is *one single 'object'* (to eliminate distinguishability), which is *extended* (to eliminate localizability) and which is *fluctuating* (to ensure approximate continuity). Nature is a far-reaching, fluctuating fold. Nature is similar to an amoeba. The tangled branches of the amoeba allow a definition of length

---

Challenge 92 ny

* Is this the only way to describe nature? Is it possible to find another description, particularly if space and time are not used as background? These questions are still subject of research.

via counting of the folds. In this way, *discreteness* of space, time, and particles could also be realized; the quantization of space-time, matter and radiation thus follows. Any flexible and deformable entity is also a perfect candidate for the realization of diffeomorphism invariance, as required by general relativity.

A simple candidate for the extended fold is the image of a fluctuating, flexible *tube*, of Planck diameter. Counting tubes implies determining distances or areas. The minimum possible count (one) gives the minimum distance, from which quantum theory is derived. In fact, at this point we can use as a model any flexible object with a small dimension, such as a tube, a thin sheet, a ball chain or a woven collection of rings. We will explore these options later on. Page 148

THE ENTROPY OF BLACK HOLES

We are still collecting arguments to determining the shape of fundamental constituents. Another approach is to study situations where particles appear in large numbers. Systems composed of many particles behave differently depending on whether the particles are point-like or extended. In particular, their entropy is different. Studying large-number entropy thus allows us to determine the shape of components. The most revealing situations are those in which large numbers of particles are crammed in a small volume. Therefore we are led to study the entropy of black holes. Indeed, black holes tell us a lot about the fundamental constituents of nature.

A black hole is a body whose gravity is so strong that even light cannot escape. It is easily deduced from general relativity that any body whose mass m fits inside the so-called Schwarzschild radius

$$r_S = 2Gm/c^2 \qquad (127)$$

is a black hole. A black hole can be formed when a whole star collapses under its own weight. Such a black hole is a macroscopic body, with a large number of constituents. Therefore it has an entropy. The entropy S of a macroscopic black hole was determined by Bekenstein and Hawking, and is given by Ref. 56, Ref. 57

$$S = \frac{k}{4l_{\mathrm{Pl}}^2}A = \frac{kc^3}{4\hbar G}A \quad \text{or} \quad S = k\frac{4\pi G m^2}{\hbar c} \qquad (128)$$

where k is the Boltzmann constant and $A = 4\pi r_S^2$ is the surface of the black hole horizon. Ref. 125 This important result has been derived in many different ways. The various derivations confirm that space-time and matter are equivalent: they show that the entropy value can be interpreted as an entropy either of matter or of space-time. In the present context, the two main points of interest are that the entropy is *finite*, and that it is *proportional to the area* of the black hole horizon.

In view of the existence of minimum lengths and times, the finiteness of the entropy is not surprising: it confirms the idea that matter is made of a finite number of discrete constituents per given volume (or area). It also shows that these constituents behave statistically: they fluctuate. In fact, quantum gravity implies a finite entropy for any object, Ref. 32 not only for black holes. Bekenstein has shown that the entropy of an object is always smaller than the entropy of a (certain type of) black hole of the same mass.

The entropy of a black hole is proportional to its horizon area. Why? This question has been the topic of a stream of publications.* A simple way to understand the entropy–surface proportionality is to look for other systems in nature whose entropy is proportional to system surface instead of system volume. In general, the entropy of a collection of flexible one-dimensional objects, such as polymer chains, shares this property. Indeed, the entropy of a polymer chain made of N monomers, each of length a, whose ends are kept a distance r apart, is given by
Ref. 126 Ref. 128 Ref. 129

$$S(r) = k\frac{3r^2}{2Na^2} \quad \text{for} \quad Na \gg \sqrt{Na} \gg r \,. \tag{129}$$

This formula can be derived in a few lines from the properties of a random walk on a lattice, using only two assumptions: the chains are extended; and they have a characteristic internal length a given by the smallest straight segment. Expression (129) is only valid if the polymers are effectively infinite: in other words, if the length Na of the chain and the *elongation* $a\sqrt{N}$, are much larger than the radius r of the region of interest. If the chain length is comparable to or smaller than the region of interest, one gets the usual extensive entropy, satisfying $S \sim r^3$. Thus *only flexible extended constituents yield an* $S \sim r^2$ *dependence.*

However, there is a difficulty. From the expression for the entropy of a black hole we deduce that the elongation $a\sqrt{N}$ is given by $a\sqrt{N} \approx l_{\mathrm{Pl}}$; thus it is much smaller than the radius of a general macroscopic black hole, which can have a diameter of several kilometres. On the other hand, the formula for long constituents is only valid when the chains are longer than the distance r between the end points.

This difficulty disappears when we remember that space near a black hole is strongly curved. All lengths have to be measured in the same coordinate system. It is well known that for an outside observer, any object of finite size falling into a black hole seems to cover the complete horizon for long times (whereas for an observer attached to the object it falls into the hole in its original size). In short, an extended constituent can have a proper length of Planck size but still, when seen by an outside observer, be as long as the horizon of the black hole. We thus find that black holes are made of extended constituents.
Ref. 114

Another viewpoint can confirm this result. Entropy is (proportional to) the number of yes-or-no questions needed to know the exact state of the system. This view of black holes has been introduced by Gerard 't Hooft. But if a system is defined by its surface, as a black hole is, its components must be extended.
Ref. 101

Finally, imagining black holes as made of extended constituents is also consistent with the so-called *no-hair theorem*: black holes' properties do not depend on what falls into them, as all matter and radiation particles are made of the same extended components. The final state only depends on the number of constituents.

Ref. 127 * The result can be derived from quantum statistics alone. However, this derivation does not yield the proportionality coefficient.

SUMMARY OF THE FIFTH ARGUMENT FOR EXTENSION

Black hole entropy is best understood as resulting from extended constituents that tangle and fluctuate. Black hole entropy also confirms that vacuum and particles are made of common constituents.

EXCHANGING SPACE POINTS OR PARTICLES AT PLANCK SCALES

Let us now focus on the exchange behaviour of fundamental constituents in nature. We saw above that 'points' in space have to be abandoned in favour of continuous, fluctuating constituents common to space, time and matter. Is such a constituent a boson or a fermion? If we exchange two points of empty space, in everyday life, nothing happens. Indeed, at the basis of quantum field theory is the relation

$$[x, y] = xy - yx = 0 \tag{130}$$

between any two points with coordinates x and y, making them bosons. But at Planck scales, because of the existence of minimal distances and areas, this relation must at least be changed to

$$[x, y] = l_{\mathrm{Pl}}^2 + \ldots \,. \tag{131}$$

This means that 'points' are neither bosons nor fermions.* 'Points' have more complex exchange properties. In fact, the term on the right-hand side will be energy-dependent, to an increasing extent as we approach Planck scales. In particular, as we have seen, gravity implies that a double exchange does not lead back to the original situation at Planck scales. Ref. 44

Constituents obeying this or similar relations have been studied in mathematics for many decades: they are called braids. Thus at Planck scales space is not made of points, but of braids or their generalizations. We find again that quantum theory and general relativity taken together imply that the vacuum must be made of extended constituents. Ref. 130 Ref. 130

We now turn to particles. All particles in nature behave in a similar way: we know that at low, everyday energies, particles of the same type are *identical*. Experiments sensitive to quantum effects show that there is no way to distinguish them: any system of several identical particles has permutation symmetry. On the other hand, we know that at Planck energy all low-energy symmetries disappear. We also know that at Planck energy permutation cannot be carried out, as it implies exchanging positions of two particles. At Planck energy, nothing can be distinguished from vacuum; thus no two entities can be shown to have identical properties. Indeed, no two particles can be shown to be indistinguishable, as they cannot even be shown to be separate.

What happens when we slowly approach Planck energy? At everyday energies, permutation symmetry is defined by commutation or anticommutation relations between particle creation operators

$$a^\dagger b^\dagger \pm b^\dagger a^\dagger = 0 \,. \tag{132}$$

---

* The same reasoning applies to the fermionic or Grassmann coordinates used in supersymmetry.

At Planck energy this cannot be correct. Quantum gravity effects modify the right-hand side: they add an energy-dependent term, which is negligible at experimentally accessible energies but which becomes important at Planck energy. We know from our experience with Planck scales that, in contrast to everyday life, exchanging particles twice cannot lead back to the original situation. A double exchange at Planck energy cannot have no effect, because at Planck energy such statements are impossible. The simplest extension of the commutation relation (132) for which the right-hand side does not vanish is *braid symmetry*. This again suggests that particles are made of extended constituents.

Ref. 44

Ref. 130

SUMMARY OF THE SIXTH ARGUMENT FOR EXTENSION

Extrapolating both point and particle indistinguishability to Planck scales suggests extended, braid-like constituents.

THE MEANING OF SPIN

As last argument we will now show that the extension of particles makes sense even at everyday energy. Any particle is a part of the universe. A part is something that is different from anything else. Being 'different' means that exchange has some effect. *Distinction means possibility of exchange.* In other words, any part of the universe is also described by its *exchange behaviour*.

In nature, exchange is composed of rotations. In other words, parts of nature are described by their *rotation behaviour*. This is why, for microscopic particles, exchange behaviour is specified by spin. *Spin distinguishes particles from vacuum.*[*]

We note that volume does not distinguish vacuum from particles; neither does rest mass or charge: nature provides particles without measurable volume, rest mass or charge, such as photons. The only observables that distinguish particles from vacuum are spin and momentum. In fact, linear momentum is only a limiting case of angular momentum. We thus find again that rotation behaviour is the basic aspect distinguishing particles from vacuum.

If spin is the central property that distinguishes particles from vacuum, finding a model for spin is of central importance. But we do not have to search for long. A model for spin 1/2 is part of physics folklore since almost a century. Any belt provides an example, as we discussed in detail when exploring permutation symmetry. Any localized structure with any number of tails attached to it – tails that reach the border of the region of space under consideration – has the same properties as a spin 1/2 particle. The only condition is that the tails themselves are *unobservable*. It is a famous exercise to show that such a model, shown in Figure 11, is indeed invariant under 4π rotations but not under 2π rotations, and that two such particles get entangled when exchanged, but get untangled when exchanged twice. Such a *tail model* has all the properties of spin 1/2 particles, independently of the precise structure of the central region, which is not important at

Vol. IV, page 94

---

[*] With a flat (or other) background, it is possible to define a *local* energy–momentum tensor. Thus particles can be defined. Without a background, this is not possible, and only global quantities can be defined. Without a background, even particles cannot be defined. Therefore, in this section we assume that we have a slowly varying space-time background.

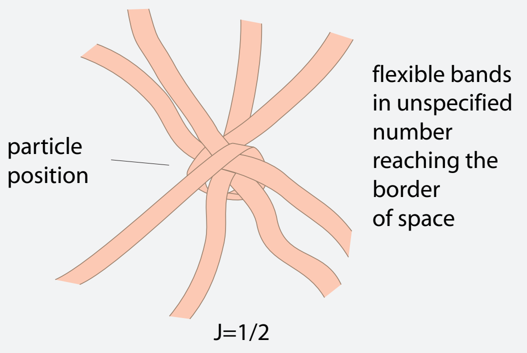


FIGURE 11 A possible model for a spin 1/2 particle.

this point. The tail model even has the same problems with highly curved space as real spin 1/2 particles have. We will explore the issues in more detail shortly. Page 166

The tail model thus confirms that rotation is partial exchange. More interestingly, it shows that rotation implies connection with the border of space-time. Ref. 131 Extended particles can be rotating. Particles can have spin 1/2 provided that they have tails going to the border of space-time. If the tails do not reach the border, the model does not work. Spin 1/2 thus even seems to *require* extension.

SUMMARY OF THE SEVENTH ARGUMENT FOR EXTENSION

Exploring the properties of particle spin suggests extended constituents in elementary fermions.

CURIOSITIES AND FUN CHALLENGES ABOUT EXTENSION

> “No problem is so formidable that you can’t walk away from it.”
>
> Charles Schultz

In case that this section has not provided enough food for thought, here is some more.

* *

Quantum theory implies that even if tight walls exist, the lid of a box made of them could never be tightly shut. Can you provide the argument? Challenge 93 s

* *

Challenge 94 e Can you provide an argument against the idea of extended constituents in nature? If so, publish it!

* *

Does duality imply that the cosmic background fluctuations (at the origin of galaxies and clusters) are the same as vacuum fluctuations? Challenge 95 ny

* *

Does duality imply that a system with two small masses colliding is the same as a system with two large masses gravitating? Challenge 96 ny

* *

It seems that in all arguments so far we have assumed that time is continuous, even though we know it is not. Does this change the conclusions? Challenge 97 d

* *

Duality also implies that in some sense large and small masses are equivalent. A mass m in a radius r is equivalent to a mass $m_{\rm Pl}^2/m$ in a radius $l_{\rm Pl}^2/r$. In other words, duality transforms mass density from ρ to $\rho_{\rm Pl}^2/\rho$. Vacuum and maximum density are equivalent. Vacuum is thus dual to black holes.

* *

Total symmetry and space-time duality together imply that there is a symmetry between all values an observable can take. Do you agree? Challenge 98 s

* *

Can supersymmetry be an aspect or special case of total symmetry, or is it something else? Challenge 99 s

* *

Any description is a mapping from nature to mathematics, i.e., from observed differences (and relations) to thought differences (and relations). How can we do this accurately, if differences are only approximate? Is this the end of physics? Challenge 100 s

* *

Duality implies that the notion of initial conditions for the big bang makes no sense, as we saw earlier by considering the minimal distance. As duality implies a symmetry between large and small energies, the big bang itself becomes a vague concept. What else do extended constituents imply for the big bang? Challenge 101 d

* *

Can you show that going to high energies or selecting a Planck-size region of space-time is equivalent to visiting the big bang? Challenge 102 d

* *

We need a description for the expansion of the universe in terms of extended constituents. Ref. 134, Ref. 147 Various approaches are being explored. Can you speculate about the solution? Challenge 103 ny

Sexual preferences in physics

Why has extension appeared so late in the history of physics? Here is a *not too serious* answer. When we discussed the definition of nature as made of tiny balls moving in a vacuum, we described this as a typically male idea. Vol. I, page 270 This implies that the female part is missing. Which part would that be?

From a general point of view, the female part of physics might be the quantum description of the vacuum. We can speculate that if women had developed physics, the order of its discoveries might have been different. Instead of studying matter first, as men did, women might have studied the vacuum first, the container of all things. And women might not have needed 2500 years to understand that nature is not made of void and little balls, but that everything in nature is made of extended constituents. It is a shame for (male) physics that this discovery took so long.

CHECKS OF EXTENSION

Is nature really described by extended constituents? The idea is taken for granted in all current research approaches to unification. How can we be sure that extension is correct? The arguments presented above provide several possible checks. We start with some options for *theoretical* falsification.

— A single Gedanken experiment invalidating extended constituents would prove extension wrong. Challenge 104 e

— Extended constituents must appear if we start from *any* physical (low-energy) concept – not only from length measurements – and study how the concept behaves at Planck scales.

— Invalidating the requirement of extremal identity would invalidate the need for extended constituents. As Edward Witten liked to say, any unified model of nature must include duality. Page 105

— Any explanation of black hole entropy without extended constituents would invalidate the need for extended constituents.

— If the measurement of length could be shown not to be related to the counting of folds of extended constituents, extension would become unnecessary.

— Finding any property of nature not consistent with extended constituents would spell the end of extension.

Any of these options, by the way, would signal the end for almost all current unification attempts. But physics is an experimental science. What kind of *data* could falsify the idea of extended constituents?

— Observing a single particle in cosmic rays with energy above the corrected Planck energy would invalidate the invariant limits and thus also extension. However, the present particle energy record, about 0.35 ZeV, is a million times lower than the Planck energy. Ref. 132

— Finding an elementary particle of spin 0 would invalidate extension. In particular, finding the Higgs boson and showing that it is elementary, i.e., that its size is smaller than its own Compton wavelength, would invalidate the model. We come back to this issue later. Page 286

— Any proposed experimental check of string theory also yields information on the idea of extended constituents. For example, Paul Mende has proposed a number of checks on the motion of extended objects in space-time. He argues that an extended object Ref. 133

and a mass point move differently; the differences could be noticeable in scattering or dispersion of light near masses.

— In 2002, the Italian physicist Andrea Gregori made a surprising prediction for any model using extended constituents that reach the border of the universe: if particles are extended in this way, their mass should depend on the size of the universe. Thus particle masses should change with time, especially around the big bang. We will discuss this conjecture later on. Ref. 134 Page 311

Experimental falsification of extension has not yet occurred. In fact, experimental falsification is quite difficult. It seems easier and more productive to *confirm* extension. Confirmation is a well-defined project: it implies to deduce all those aspects of nature that are given in the millennium list of unexplained properties. Among others, confirmation requires to find a concrete model, based on extended constituents, for the electron, the muon, the tau, and their neutrinos. Confirmation also requires using extended constituents to realize an old dream of particle physics: to deduce the value of the coupling constants and particle masses. Page 18

CURRENT RESEARCH BASED ON EXTENDED CONSTITUENTS

> “To understand is to perceive patterns.”
> Isaiah Berlin

The Greeks deduced the existence of atoms from the fact that fish can swim through water. They argued that only if water is made of atoms could a fish make its way through it, by pushing the atoms aside. We can ask a similar question of a particle flying through a vacuum: why is it able to do so? A vacuum cannot be a fluid or a solid composed of small constituents, as its dimensionality would not then be fixed. Only one possibility remains: both vacuum and particles are made of a web of extended constituents. Ref. 118

The idea of describing matter as composed of extended constituents dates from the 1960s. That of describing nature as composed of ‘infinitely’ extended constituents dates from the 1980s. In addition to the arguments presented so far, current research provides several other approaches that arrive at the same conclusion.

* *

Bosonization, the construction of fermions using an infinite number of bosons, is a central aspect of modern unification attempts. It also implies coupling duality, and thus the extension of fundamental constituents. Ref. 135

* *

Ref. 137, Ref. 138 Research into quantum gravity – in particular the study of spin networks, spin foams and loop quantum gravity – has shown that the vacuum can be thought of as a collection of extended constituents.

* *

In the 1990s, Dirk Kreimer showed that high-order QED Feynman diagrams are related to knot theory. He thus proved that extension arrives by the back door even when electromagnetism is described in terms of point particles. Ref. 139

* *

Ref. 140 A popular topic in particle physics, 'holography', relates the surface and the volume of physical systems at high energy. It implies extended constituents of nature.

* *

Vol. IV, page 132 It is long known that wave function collapse can be seen as the result of extended constituents. We will explore the details below.

* *

Ref. 141, Ref. 142 Ref. 143, Ref. 144 Ref. 145, Ref. 146 At the start of the twenty-first century a number of new promising approaches to describe elementary particles appeared, such as models based on string nets, models based on bands, models based on ribbons, and models based on knots. All these attempts make use of extended constituents. We now look more closely at an attempt that was especially popular between the years 1984 and 2005.

Superstrings – extension and a web of dualities

> Throw physic to the dogs; I'll none of it.
>
> William Shakespeare, *Macbeth*.

Ref. 149 *Superstrings* and *supermembranes* – often simply called *strings* and *membranes* – are extended constituents in the most investigated physics conjecture ever. The approach contains a maximum speed, a minimum action and a maximum force (or tension). The approach thus incorporates special relativity, quantum theory and general relativity. This attempt to achieve the final description of nature uses four ideas that go beyond standard general relativity and quantum theory:

1. Particles are conjectured to be *extended*. Originally, particles were conjectured to be one-dimensional *oscillating strings*. In a subsequent generalization, particles are conjectured to be fluctuating higher-dimensional membranes.
2. The conjecture uses *higher dimensions* to unify interactions. A number of space-time dimensions much higher than 3+1, typically 10 or 11, is necessary for a mathematically consistent description of strings and membranes.
3. The conjecture is based on *supersymmetry*. Supersymmetry is a symmetry that relates matter to radiation, or equivalently, fermions to bosons. Supersymmetry is the most general local interaction symmetry that can be constructed mathematically. Supersymmetry is the reason for the terms 'superstring' and 'supermembrane'.
4. Ref. 150 The conjecture makes heavy use of *dualities*. In the context of high-energy physics, dualities are symmetries between large and small values of physical observables. Important examples are space-time duality and coupling constant duality. Dualities are global interaction and space-time symmetries. They are essential for the inclusion of gauge interaction and gravitation in the quantum description of nature. Dualities also express a fundamental equivalence between space-time and matter–radiation. Dualities also imply and contain *holography*, the idea that physical systems are completely fixed by the states on their bounding surface.

By incorporating these four ideas, the *string conjecture* – named so by Brian Greene, one of its most important researchers – acquires a number of appealing characteristics. Ref. 151

Why superstrings and supermembranes are so appealing

First of all, the string conjecture is unique: the Lagrangian is unique and has no adjustable parameters. Furthermore, as we would expect from a description involving extended constituents, the conjecture includes gravity. In addition, the conjecture describes interactions: it describes gauge fields. The conjecture thus expands quantum field theory, while retaining all its essential points. In this way, the conjecture fulfils most of the requirements for a unified description of motion that we have deduced so far. For example, particles are not point-like, there are minimal length and time intervals, and all other limit quantities appear. (Only the requirement of the lack of sets seems not fulfilled.)

The string conjecture has many large symmetries, which arise from its many dualities. These symmetries connect many situations that seem intuitively to be radically different: this makes the conjecture extremely fascinating, but also difficult to picture.

The conjecture shows special cancellations of anomalies and other inconsistencies. Historically, the first example was the Green–Schwarz anomaly cancellation; but strings also solve many other mathematical problems of quantum field theory.

Edward Witten, the central figure of the field, likes to say that quantum theory cures the infinities that appear in e^2/r when the distance r goes to zero; in the same way, strings cure the infinities that appear in m^2/r when the distance r goes to zero.

Also following Witten, in the string conjecture, the interactions follow from the particle definitions: interactions do not have to be added. That is why the string conjecture is so powerful. In particular, it predicts gravity, gauge theory, supersymmetry and supergravity.

About gravity, one of the pretty results of the string conjecture is that strings and black holes are complementary to each other. This was argued by Polchinsky, Horowitz and Susskind. Ref. 152

As expected, strings explain the entropy of black holes. Strominger and Vafa showed this important result in 1996. Ref. 125

The string conjecture naturally includes *holography*, the idea that the degrees of freedom of a physical system are determined by its boundary. In particular, holography provides for a deep duality between gauge theory and gravity. More precisely, there is a correspondence between quantum field theory in flat space and string theory in certain higher-dimensional spaces that contain anti-de Sitter space.

In short, the string conjecture implies fascinating mathematics. Conformal invariance enters the Lagrangian. Concepts such as the Virasoro algebra, conformal field theory, topological field theory and many related ideas provide vast and fascinating generalizations of quantum field theory.

Why the mathematics of strings is so difficult

The string conjecture, like all modern descriptions of physics, is described by a Lagrangian. The Lagrangian is constructed starting from the Lagrangian for the motion of a classical string of matter. Then the Lagrangian for the corresponding quantum string fields is constructed, and then higher dimensions, supersymmetry, dualities and membranes are incorporated. This formulation of the string conjecture takes for granted the existence of a space-time background.

The Lagrangian of the string conjecture is extremely complex, much too complex to write it down here. It is not as simple as the Lagrangian of the standard model of particle physics or the Lagrangian of general relativity. But the complexity of the Lagrangian is not the only reason why the studying the string conjecture is difficult.

It turns out that exploring how the known 4 dimensions of space-time are embedded in the 10 or 11 dimensions of the string conjecture is extremely involved. The topology and the size of the additional dimensions is unclear. There are only few people who are able to study these options.

TESTING STRINGS: COUPLINGS AND MASSES

One of the main results of quantum chromodynamics or QCD, the theory of strong interactions, is the explanation of mass relations such as Ref. 153

$$m_{\text{proton}} \sim \mathrm{e}^{-k/\alpha_{\text{Pl}}} m_{\text{Pl}} \quad \text{and} \quad k = 11/2\pi \;, \; \alpha_{\text{Pl}} \approx 1/25 \;. \tag{133}$$

Here, the value of the strong coupling constant α_{Pl} is taken at the Planck energy. In other words, a general understanding of masses of bound states of the strong interaction, such as the proton, requires little more than a knowledge of the Planck energy and the coupling constant at that energy. The approximate value $\alpha_{\text{Pl}} = 1/25$ is an empirical value based on experimental data. Page 318

Any unified theory must allow us to calculate the three gauge coupling constants as a function of energy, thus including the value α_{Pl} at Planck energy. At present, most researchers regard the search for the vacuum state – the precise embedding of four dimensions in the total ten – as the main difficulty facing the string conjecture. Without knowledge of the vacuum state, no calculations of coupling constants or masses are possible.

The vacuum state of the string conjecture is expected to be one of an extremely involved set of topologically distinct manifolds. It is estimated that there are around 10^{500} candidate vacuum states. The universe contains 10^{80} atoms; it thus seems easier to find a particular atom in the universe that to find the correct vacuum state. All the advantages that are due to a unique Lagrangian are lost again. Ref. 154

We can also describe the problems with the calculation of masses in the following way. The string conjecture predicts states with Planck mass and with zero mass. The zero-mass particles are then thought to get their actual mass, which is tiny compared with the Planck mass, from the Higgs mechanism. However, the Higgs mechanism and its parameters have not yet been deduced from strings.

The status of strings

> Es ist nicht Großes ohne Leidenschaft
> vollbracht worden, noch kann es ohne solche
> vollbracht werden.*
> Friedrich Hegel, *Enzyklopädie*.

Historically, the community of string researchers took over 10 years to understand that strings were *not* the basic entities of the string conjecture. The fundamental entities are *membranes*. After another 10 years, and more, it became unclear whether membranes are the most practical fundamental entities for calculation. The search for the most practical entities is still ongoing.

It is estimated that over 10 000 man-years have been invested in the exploration of the string conjecture: compare this with about a dozen man-years for electrodynamics, a dozen man-years for general relativity, and a dozen man-years for the foundation of quantum theory.

In fact, it is fair to say that nowadays, string researchers are stuck. Despite over 10 000 man-years of effort, not a single calculation of an *experimentally measurable* value has been made. For example, the string conjecture has not predicted the masses of any elementary particle, nor the value of any coupling constant, nor the number of gauge interactions. Worse, *none* of the open issues from the millennium list has been solved by the string conjecture. This disappointing situation is the reason why many scholars, including several Nobel Prize winners, dismiss the string conjecture altogether.

Page 18

Ref. 155

What are the reasons that the string conjecture, like several other approaches based on extended constituents, was unsuccessful? Superstrings and supermembranes, like many other proposed fundamental constituents, are *complex* structures in themselves: strings and membranes move in many dimensions, carry mass, have tension and carry fields. In fact, the precise mathematical definition of a string or a membrane and their features is so complex that already understanding the definition is beyond the capabilities of most physicists. But a high complexity always nourishes the doubt that some of the underlying assumptions do not apply to nature.

Put in different terms, the superstring conjecture was not successful because its *basic principles* have never been clarified. This lack of clear foundations is regularly underlined even by supporters of the string conjecture, such as Murray Gell-Mann. Despite this gap, no research papers on the basic principles exist to this day.

Summary on extension in nature

> Wir müssen wissen, wir werden wissen.**
> David Hilbert, 1930

We have explored topics such as the Planck limits, three-dimensionality, curvature, particle shape, renormalization, spin, bosonization, the cosmological constant problem, and

---

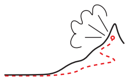
* 'Nothing great has been achieved without passion, nor can it be achieved without it.' Hegel writes this towards the end of the third and last part of his *Enzyklopädie der philosophischen Wissenschaften im Grundrisse*, §474, 296.

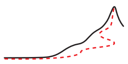
** 'We must know, we will know.' This was Hilbert's famous personal credo.

Ref. 148 the search for a 'background-free' description of nature. We have seen that at Planck scales, all explorations lead to the same conclusion: what we usually call space-time points and point particles are in fact made of *extended* constituents.

Despite using extension as fundamental aspect, and despite many interesting results, no attempt from the twentieth century, not even the string conjecture, has been successful. A different approach to calculations with extended constituents is required: an approach that is based on clear principles and that has few assumptions.

In our quest for a final theory of physics, we can advance by posing the following problem. The most explored approach from the twentieth century, the string conjecture, is based on four basic assumptions: extension, duality, higher dimensions and supersymmetry. Can we dispense with any of them? Now, duality is closely related to extension, for which enough theoretical and experimental evidence exists, as we have argued above. On the other hand, the expressions for the Schwarzschild radius and for the Compton Page 66, page 72 wavelength imply, as we found out above, that the dimensionality of space and the statistics of particles are undefined at Planck scales. In other words, nature does not have higher dimensions nor supersymmetry at Planck scales. Thus we drop these two incorrect assumptions and continue our adventure.





CHAPTER 8

THE BASIS OF THE STRAND MODEL

> We haven't got the money, so we have to think.
>
> Ernest Rutherford

The two extremely precise descriptions of nature that were discovered in the twentieth century – quantum field theory and general relativity – are the low-energy approximations of physics at the Planck scale. To find the final and unified description that is valid at the Planck scale itself, we follow the method that has been the most effective during the history of physics: we search for the *simplest* possible description. Simplicity was used successfully, for example, in the discovery of special relativity, in the discovery of quantum theory, and in the discovery of general relativity. We therefore use the guidance provided by simplicity to deduce the final and unified description of motion.

The central requirement for any *unified* description is that it leads from Planck scales, and thus from Planck units, to quantum field theory, to the standard model of elementary particles and to general relativity. In simple terms, as discussed below, the unified description must be valid for *all observations* and provide *complete precision*.

Page 158

From the preceding chapters, we know already quite a bit about the unified description. In particular, any unified description of general relativity and quantum theory must use *extended* constituents. We discovered a number of reasons that are central for this conclusion. All these reasons appear only when quantum theory and general relativity are combined. First, only constituents that are extended allow to deduce black hole entropy. Secondly, only extended constituents allow to model that elementary particles are not point-like or that physical space is not made of points. Thirdly, only extended constituents allow to model a smallest measurable space and time interval. Fourthly, only extended constituents allow to model spin 1/2.

But we are not only looking for a unified theory; we are also looking for the *final* theory. This implies a second requirement: the final theory must be *unmodifiable*. As we will argue below, if a candidate description can be modified, or generalized, or reduced to special cases, or varied in any other way, it is not final.

Page 159

REQUIREMENTS FOR A FINAL THEORY

In our quest for the final, unified theory, we have deduced many requirements that such a theory must realize. All the requirements that we deduced are listed in Table 7. So far, the table is not found elsewhere in the literature. Certain requirements follow from

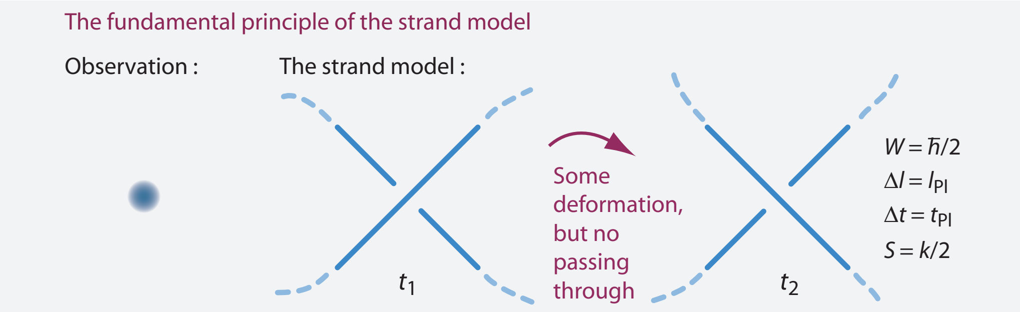


FIGURE 12 The fundamental principle of the strand model: the simplest observation in nature, a 'point-like' event, is defined by a crossing switch in three spatial dimensions. The crossing switch defines the action $\hbar/2$, the Planck length, the Planck time and the Boltzmann constant.

Challenge 105 e

the property that the description must be final, others from the property that it must be unified, and still others from the property that it must describe nature. More specifically, the technical requirements all follow when the expressions for the Compton wavelength and for the Schwarzschild radius are combined.

TABLE 7 General requirements for a final and unified description of nature and of motion.

ASPECT	REQUIREMENTS FOR THE FINAL AND UNIFIED DESCRIPTION
Precision	must be *complete*; the unified description must describe all motion and explain all open issues from the millennium list, given in Table 9 on page 158.
Modification	must be *impossible*, as explained on page 159.
Fundamental principles	must be *clear*. (Otherwise the unified description is not falsifiable.)
Vacuum and particles	*must not differ* at Planck scales because of limits of measurement precision; vacuum and particles thus must be described by *common* fundamental constituents.
Fundamental constituents	must be *extended* and *fluctuating*, to explain black hole entropy, spin, space-time homogeneity and isotropy of space.
Fundamental constituents	must be as *simple* as possible, to satisfy Occam's razor.
Fundamental constituents	must determine all observables.
Fundamental constituents	must be the *only unobservable* entities. (If they were observable, the theory would not be final; if more unobservable entities would exist, the theory would be fiction, not science.)
Non-locality	must be part of the description; non-locality must be negligible at everyday scales, but important at Planck scales.
Physical points and sets	must *not exist*, due to limits of measurement precision; points and sets only exist approximately, at everyday scales.
Evolution equations	must *not exist*, due to the lack of points and sets.
Physical systems	must not exist at Planck scales, due to limits of measurement precision; systems only exist approximately at everyday scales.
Universe	must not be a system, due to limits of measurement precision.

TABLE 7 (Continued) General requirements for a final and unified description of nature and of motion.

Aspect	Requirements for the final and unified description
Planck's natural units	must be *limit values* for each observable (within a factor of order one); infinitely large or small measurement values must not exist.
Planck scale description	must imply quantum field theory, the standard model of particle physics, general relativity and cosmology.
Planck's natural units	must define all observables, including coupling constants.
Relation to experiment	must be as simple as possible, to satisfy Occam's razor.
Background dependence	is required, as background independence is logically impossible.
Background space-time	must be *equal* to physical space-time at everyday scale, but must *differ* globally and at Planck scale.
Big bang	must not be an event, and thus not be a beginning, as this would contradict the non-existence of points and sets in nature.
Circularity of definitions	of physical concepts must be part of the final, unified description, as a consequence of being 'precise talk about nature'.
Axiomatic description	must be impossible, as nature is not described by sets; Hilbert's sixth problem must have no solution.
Dimensionality of space	must be *undefined* at Planck scale, as space is undefined there.
Symmetries	must be *undefined* at Planck scale, due to the limits to measurement precision.
Large and small scales	must be *similar*, due to the limits to measurement precision.

Looking at the table of requirements for a final theory, we note something astonishing. Even though all requirements appear when quantum physics and general relativity are combined, each of these requirements *contradicts* both quantum physics and general relativity. The final theory thus *differs* from both pillars of modern physics. A final theory cannot be found if we remain prisoners of either quantum theory or general relativity. To put it bluntly, each requirement for the final theory contradicts every result of twentieth century physics.

Page 160

Of the few candidate descriptions that satisfy the requirements of the table, it seems that the *simplest* is the one that uses *featureless fluctuating strands*. In this approach, strands,* not points, are assumed to be the fundamental constituents of vacuum, matter and radiation.

INTRODUCING STRANDS

To describe observations, the strand model uses only one basic postulate or *fundamental principle*:

▷ ***Planck units*** are defined through crossing switches of strands.

Alternatively,

---

* In Dutch: draden, in French: fils, in German: Fäden, in Italian: fili.

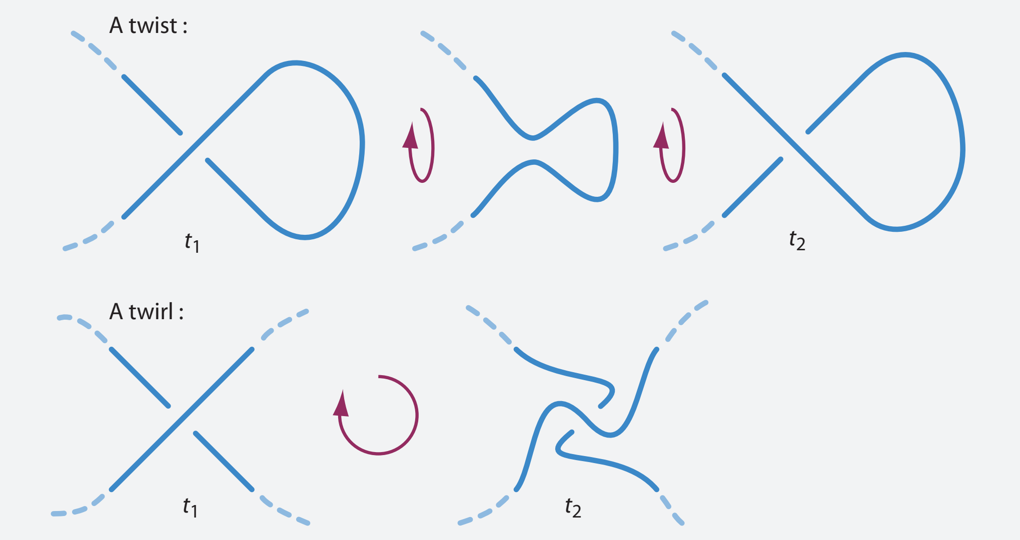


FIGURE 13 An example of strand deformation leading to a crossing switch (above) and one that does not lead to a crossing switch (below).

▷ An ***event*** is the switch of a crossing between two strand segments.

The definitions of Planck units and events as basic front-to-back switches are illustrated in Figure 12. Note that strands are *impenetrable*; realizing a crossing switch thus always requires the motion of strand segments *around* each other. A simple example of a deformation leading to a crossing switch is shown in Figure 13.

The strand model states that every event in nature is characterized by the (corrected) Planck time $t_{\mathrm{Pl}} = \sqrt{4G\hbar/c^5}$, the (corrected) Planck length $l_{\mathrm{Pl}} = \sqrt{4G\hbar/c^3}$, the Planck entropy, i.e., the Boltzmann constant k, and Planck's quantum of action $\hbar$. More precisely, the crossing switch shown in Figure 12 corresponds to an action $\hbar/2$, while $\hbar$ corresponds to a full turn.

From strands to modern physics

Challenge 106 e A quick check shows that the fundamental principle implies that the Planck units are observer-invariant limit values. Therefore, the fundamental principle contains quantum theory, thermodynamics, special and general relativity. In theory, this argument is sufficient to show that the fundamental principle contains all these parts of twentieth century physics. (We note that the standard model of particle physics is *not* found in this way; this turns out to be more involved.) In practice, physicists do not change their thinking habits that quickly; thus we show this result in more detail now.

We will discover that crossing switches define all continuous observables of modern physics. To start, we note that crossing switches are automatic consequences of the shape fluctuations of strands. We will then show that all the continuous quantities we are used to – physical space, physical time, gauge fields and wave functions – result from *averaging* crossing switches. The main conceptual tools necessary in the following are:

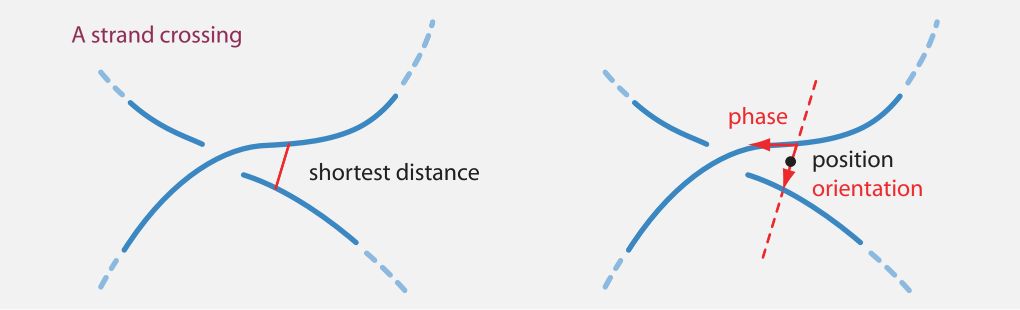


FIGURE 14 The definition of a crossing, its position, its orientation, and its phase.

- ▷ A ***crossing*** of strands is a local minimum of strand distance. The position, orientation and the phase of a crossing are defined by the space vector corresponding to the local minimum, as shown in Figure 14. The sign of the orientation is defined by arbitrarily selecting one strand as the starting strand. The arbitraryness in the definition of the phase will be of great importance later on.

- ▷ A ***crossing switch*** is the rotation of the crossing orientation by an angle π. More precisely, a crossing switch is the *inversion* of the orientation at a specific position.

We note that the definitions make use of all *three* dimensions of space; therefore the number of crossings and of crossing switches is *independent* of the direction of observation. This contrasts with the definition of crossing used in two-dimensional knot diagrams; in two-dimensional projections, the number of crossings does depend on the direction of the projection.

We also note that strand fluctuations do not conserve the number of crossings; crossings can disappear and appear over time.

The fundamental principle declares that events are not points on manifolds; instead, *events are observable crossing switches of unobservable strands.* Since all observations are events, all experimental observations should follow from the strand definition of an event. We will confirm this in the rest of this text. In other words, the strand model asserts:

Page 304

- ▷ ***Nature*** is built from fluctuating featureless strands.

The strands are featureless: they have no mass, no tension, no stiffness, no branches, no fixed length, no diameter, no ends, and they cannot be cut or pushed through each other. Strands have no measurable property at all: strands are unobservable. Only crossing switches are observable. Featureless strands are thus among the simplest possible extended constituents. How simple? We will discuss this shortly.

Page 160

- ▷ ***Strands*** are one-dimensional curves in three-dimensional space that reach the border of space.

TABLE 8 Correspondences between physical systems and mathematical tangles.

PHYSICAL SYSTEM	STRANDS	TANGLE TYPE
Vacuum and dark energy	many infinite unknotted strands	unlinked, trivial
Elementary vector boson (radiation)	one infinite strand	family of unknotted or knotted curves
Elementary fermion (matter)	two or three infinite strands	family of braided or rational tangles
Graviton	two infinite twisted strands	rational tangle
Gravity wave	many infinite twisted strands	many rational tangles
Horizon	many woven infinite strands	web-like tangle

In practice, the *border of space* has one of two possible meanings. Whenever space is assumed to be flat, the border of space is spatial infinity. Whenever we take into account the properties of the universe as a whole, the border of space is the cosmic horizon.

> ▷ *All* strand fluctuations are possible, as long as the strands do not interpenetrate. In particular, there is no speed limit for strands. Whenever strand fluctuations lead to a crossing switch, they lead to an observable effect.

The strand model asserts that *everything* that occurs in nature is due to fluctuating strands. Fluctuating strands explain everything that *does* happen, and explain everything that does *not* happen. Our main aim in the following is to classify all possible strand fluctuations and, by doing so, to classify everything that we observe in nature.

In particular, we will discover that *all* physical systems can be constructed from strands. Table 8 lists the general classes. In all physical systems, the shape fluctuations of tangles lead to crossing switches. We will discover that strand fluctuations lead to the usual evolution of matter, radiation and vacuum curvature. In particular, we will discover that strands and their crossing switches allow to deduce the evolution equations of quantum field theory and of general relativity.

The strand model asserts that matter, radiation and vacuum are all built from *fluctuating strands* in a *continuous background*. We will show shortly why both concepts are required. We first clarify these two basic terms.

Page 162

> ▷ ***Continuous background space*** is introduced by the observer, in order to be able to describe observations. Every observer introduces his own background. But all backgrounds have three spatial and one temporal dimension.

At this point of the discussion, we simply assume background space. Later on we will see why background space appears and why it *needs* to be three-dimensional. The *size* of the background space is assumed to be large – ideally, to be infinitely large.

Page 192

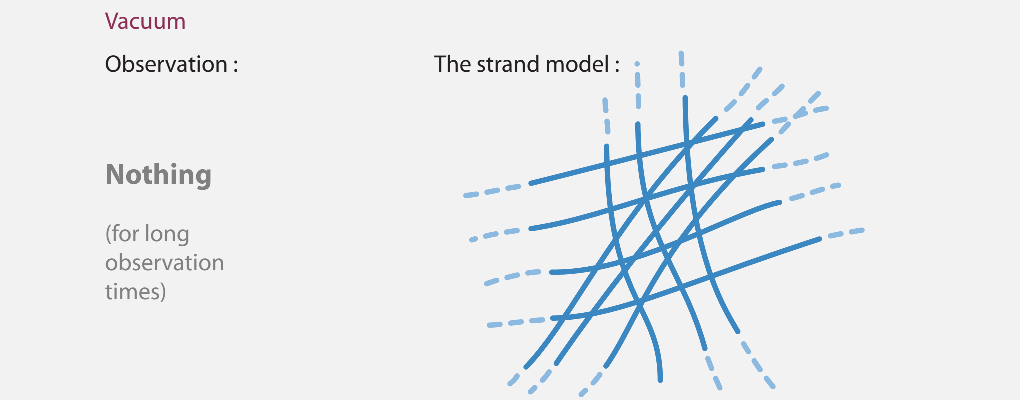


FIGURE 15 A schematic illustration of the strand model for the vacuum, or physical space.

- ▷ ***Fluctuations*** change the position and the shape of strands; fluctuations thus change curvature, length, orientation and position of strand segments. However, fluctuations never allow one strand to pass through another.

Using the idea of fluctuations we can now construct, step by step, all important physical systems. We start with the most important.

- ▷ ***Vacuum***, or ***physical space***, is formed by the time average of many unknotted fluctuating strands.

Vacuum and physical space are taken to be synonyms from now on; the exploration will show that this is the most sensible use of the two concepts.* However, the strand model distinguishes *physical* space from *background* space. In particular, since matter and vacuum are made of the same constituents, it is impossible to speak of physical space at the location of matter. At the location of matter, it is only possible to use the concept of background space.

When the strand fluctuations in flat vacuum are averaged over time, there are no crossing switches. Equivalently, if we use concepts to be introduced shorty, flat vacuum shows, averaged over time, no knots and no tangles, so that it is observed to be empty of matter and radiation. Figure 15 helps visualizing the situation. Temporary tangles that appear for a short time through vacuum fluctuations will be shown later to represent virtual particles.

Page 248

We note that the (flat) physical vacuum state, which appears after averaging the strand crossings, is *continuous*, *Lorentz invariant* and *unique*. These are important points for the consistency of the model. Later we will also discover that curvature and horizons have a natural description in terms of strands; exploring them will yield the field equations of general relativity.

With the definition of the vacuum as a time average, the strand model yields a minimum length and a continuous vacuum at the same time. In this way, many issues about

---

* We recall that since over a century, the concept of aether is superfluous, because it is synonymous with the concept of vacuum.

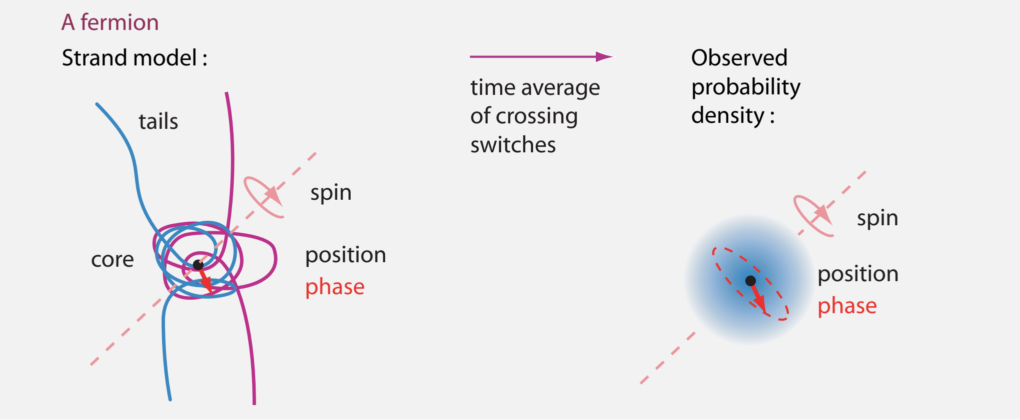


FIGURE 16 The tangle model of a spin 1/2 particle.

the alleged contradiction between continuity and minimum length are put to rest.

OBSERVABLES

The fundamental principle relates crossing switches and observations. The fundamental principle was discovered because it appears to be the only simple definition of Planck units that on one hand yields space-time continuity, isotropy and deformation, and on the other hand realizes the known connection between the quantum of action, spin and rotation.

First of all, the fundamental principle implies:

- ▷ The ***distance*** between two particles is the maximum number of crossing switches that could appear between them. Length measurement is thus defined as counting Planck lengths.
- ▷ The ***time interval*** between two events is the maximum number of crossing switches that could appear between them. Time measurement is thus defined as counting Planck times.
- ▷ The physical ***action*** of a physical system evolving from an initial to a final state is the number of crossing switches that take place. Action measurement is thus defined as counting crossing switches. Physical action is thus a measure for the *change* that a system undergoes.
- ▷ The ***entropy*** of any physical system is related to the total number of crossing switches that are possible. Entropy measurement is thus defined through the counting of crossing switches. The strand model thus states that *any large physical system* – be it made of matter, radiation, empty space or horizons – has entropy.

It is well-known that all other physical observables are defined using these four basic ones. In other words, *all* physical observables are defined with crossing switches. We also note that even though counting always yields an integer, the result of a physical measurement is often an *average* of many counting processes. As a result of averaging and fluctuations, measured values can be non-integer multiples of Planck units. Therefore, space, time, action, entropy and all other observables are effectively real numbers. Continuity is thus reconciled with the existence of a minimum measurable length and time interval. Finally, we note that the definition of observables with crossing switches automatically makes the Planck units c, $\hbar$, $c^4/4G$, k and all their combinations both *observer-invariant* and *limit* values.

Strands also define the opposite of empty space, illustrated in Figure 16:

▷ A ***quantum particle*** is a *tangle* of fluctuating strands. The tangle *core*, the region where the strands are linked or knotted, defines position, speed, phase and spin of the particle. The tangle *tails* reach up to the border of space.

Page 169 As shown in more detail soon, this definition of quantum particles yields, depending on the tangle details, either fermion or boson behaviour, reproduces the spin–statistics theorem, and defines wave functions and gauge fields. In particular, modelling fermions as tangles allows us to deduce Dirac's equation for relativistic quantum particles, as well as the interactions and the gauge symmetries of the standard model of particle physics. Still later, we will discover that tangles imply only a *finite* number of possible elementary particles, and that the *type* of tangle determines the mass, mixings, quantum numbers, charges and couplings of each elementary particle. In the 1960s, John Wheeler aimed at a description of nature that explained 'mass without mass, charge without charge, field without field'. The strand model realizes this aim.

Page 198 Page 207 Page 273

CURIOSITIES AND FUN CHALLENGES ABOUT STRANDS

There is an intuitive explanation for the central role of crossing switches in the definition of observables. All measurements – be they measurements of position, speed, mass or any other observable – are electromagnetic. In other words, all measurements in nature are detection of photons. And photon absorption and detection are intimately related to the crossing switch, as we will find out later on. Page 211

* *

Is there a limit to the fluctuations of strands? Yes and no. On one hand, the 'speed' of fluctuations is unlimited. On the other hand, fluctuations with a 'radius' smaller than a Planck length do not lead to observable effects. Note that the terms 'speed' and 'radius' are between quotation marks because they are unobservable. Care is needed when talking about strands and their fluctuations.

* *

What are strands made of? This question tests whether we are really able to maintain the fundamental circularity of the unified description. Strands are featureless. They have no measurable properties: they have no branches, carry no fields and, in particular, they cannot be divided into parts. The 'substance' that strands are made of has no properties. Thus strands are not made of anything. This may seem surprising at first. Strands are extended, and we naturally imagine them as sequence of points. But this is a fallacy.

Given the way that observations and events are defined, there is no way to observe, to label or to distinguish points on strands. Crossing switches do not allow doing so, as is Challenge 107 e easily checked: the mathematical points we imagine on a strand are not physical points. 'Points' on strands are unobservable: they simply do not exist.

But strands must be made of something, we might insist. Later we will find out that Page 266 in the strand model, the universe is made of a single strand folded in a complicated way. This strand is everything there is in nature. In short, strands are not made of something, they are made of everything. The substance of strands is nature itself.

* *

Challenge 108 e Can macroscopic determinism arise at all from randomly fluctuating strands?

* *

Challenge 109 e Do parallel strands form a crossing?

* *

Challenge 110 s Do two distant strands form a crossing?

* *

Challenge 111 s Is a crossing switch defined in more than three dimensions?

* *

Challenge 112 s Can you find a way to generalize or to modify the strand model?

* *

Is the strand model confirmed by independent research? Yes; a few years after the strand model appeared, this is starting to happen. In a long article exploring the small scale structure of space-time from various different research perspectives in general relativity, Steven Carlip comes to the conclusion that all these perspectives suggest the common Ref. 156 result that "space at a fixed time is thus threaded by rapidly fluctuating lines".

Other research approaches that confirm the strand model are mentioned in various places later in the text. Despite these developments, the essential point remains to check how the strand model compares with experiment.

Do strands unify? – The millennium list of open issues

Does the strand model reproduce all the paradoxical results we found in the first chapters? Yes, it does. The strand model implies that vacuum cannot be distinguished from Page 59 matter at Planck scales: both are made of strands. The strand model implies that observ- Page 68 ables are not real numbers at Planck scales. The strand model implies that the universe Page 85 and the vacuum are the same, when explored at high precision: both are made of one strand. The strand model also implies that the number of particles in the universe is Page 130 not clearly defined and that nature is not a set. You can check by yourself that all other Challenge 113 e paradoxes appear automatically and almost all requirements of Table 7 are fulfilled as well. Two requirements of the table must be discussed in more detail: the requirements of complete precision and of unmodifiability. We start with complete precision.

If strands really describe *all* of nature, they must explain the inverse square dependence with distance of the electrostatic and of the gravitational interaction. But that is not sufficient. If the strand model is a final, unified description, it must provide *complete* precision. First of all, the model must describe *all* experiments. This is the case, as will be shown below, because the strand model contains both general relativity and the standard model of particle physics. But above all, the model must also settle all those questions that were left unanswered by twentieth-century fundamental physics. These questions, the *millennium list* of open issues, are given (again) in Table 9.

TABLE 9 The millennium list: *everything* particle physics and general relativity *cannot* explain; thus, also the list of the *only* experimental data available to test the final, unified description of motion.

OBSERVABLE	PROPERTY UNEXPLAINED IN THE YEAR 2000
Local quantities, from quantum field theory: particle properties	
$\alpha = 1/137.036(1)$	the low energy value of the electromagnetic coupling constant
α_w or θ_w	the low energy value of the weak coupling constant or the value of the weak mixing angle
α_s	the value of the strong coupling constant at one specific energy value
m_q	the values of the 6 quark masses
m_l	the values of 6 lepton masses
m_W	the value of the mass of the W vector boson
m_H	the value of the mass of the scalar Higgs boson
$\theta_{12}, \theta_{13}, \theta_{23}$	the value of the three quark mixing angles
δ	the value of the CP violating phase for quarks
$\theta^\nu_{12}, \theta^\nu_{13}, \theta^\nu_{23}$	the value of the three neutrino mixing angles
$\delta^\nu, \alpha_1, \alpha_2$	the value of the three CP violating phases for neutrinos
$3 \cdot 4$	the number of fermion generations and of particles in each generation
J, P, C, etc.	the origin of all quantum numbers of each fermion and each boson
Local mathematical structures, from quantum field theory	
$c, \hbar, k$	the origin of the invariant Planck units of quantum field theory
$3 + 1$	the number of dimensions of physical space and time
SO(3,1)	the origin of Poincaré symmetry, i.e., of spin, position, energy, momentum
$S(n)$	the origin of particle identity, i.e., of permutation symmetry
Gauge symmetry	the origin of the gauge groups, in particular:
U(1)	the origin of the electromagnetic gauge group, i.e., of the quantization of electric charge, as well as the vanishing of magnetic charge
SU(2)	the origin of weak interaction gauge group, its breaking and P violation
SU(3)	the origin of strong interaction gauge group and its CP conservation
Ren. group	the origin of renormalization properties
$\delta W = 0$	the origin of wave functions and the least action principle in quantum theory
$W = \int L_\mathrm{SM}\,\mathrm{d}t$	the origin of the Lagrangian of the standard model of particle physics
Global quantities, from general relativity: vacuum and energy properties	
0	the observed flatness, i.e., vanishing curvature, of the universe

TABLE 9 (Continued) *Everything* the standard model and general relativity *cannot* explain.

OBSERVABLE	PROPERTY UNEXPLAINED IN THE YEAR 2000
$1.2(1) \cdot 10^{26}$ m	the distance of the horizon, i.e., the 'size' of the universe (if it makes sense)
$\rho_{\text{de}} = \Lambda c^4/(8\pi G)$ $\approx 0.5\,\text{nJ/m}^3$	the value and nature of the observed vacuum energy density, dark energy or cosmological constant
$(5 \pm 4) \cdot 10^{79}$	the number of baryons in the universe (if it makes sense), i.e., the average visible matter density in the universe
$f_0(1, ..., c.\ 10^{90})$	the initial conditions for *c.* 10^{90} particle fields in the universe (if or as long as they make sense), including the homogeneity and isotropy of matter distribution, and the density fluctuations at the origin of galaxies
ρ_{dm}	the density and nature of dark matter
Global mathematical structures, from general relativity	
c, G	the origin of the invariant Planck units of general relativity
$\delta \int L_{\text{GR}} \text{d}t = 0$	the origin the least action principle and the Lagrangian of general relativity
$\text{R} \times \text{S}^3$	the observed topology of the universe

These requirements are valid for any unified description of nature, and thus also for the strand model. They can be summarized in two general points: First, *reproduce* quantum theory, the standard model, general relativity and cosmology. Secondly, *explain* masses, mixing angles and coupling constants. Of course, only the second point is the *definite test* for a final theory. But we need the first point as well.

ARE STRANDS FINAL? – ON GENERALIZATIONS AND MODIFICATIONS

> “The chief attraction of the theory lies in its logical completeness. If a single one of the conclusions drawn from it proves wrong, it must be given up; to modify it without destroying the whole structure seems impossible.”
>
> Albert Einstein, The Times, 28. 11. 1919.

If a description of motion claims to be *final*, it must explain *all* aspects of motion. But a full explanation must be unmodifiable. This is an important point that is rarely discussed.

Theoretical and mathematical physicists are fond of *generalizing* models. If you have a description of a part of nature, they will try to find more general cases. For any candidate unified description, they will try to explore the model in more than three dimensions, with more than three generations of quarks, with more involved gauge symmetries, with different types of supersymmetry, with more Higgs doublets, or with additional heavy neutrinos. In case of the strand model, they will also explore models with more complex entities than strands, such as bands or bifurcating entities, any many more.

If a description of nature is *final*, generalizations must be *impossible*. If it were possible to generalize the unified description, it would lose the ability to explain any of the millennium issues! If a candidate unified theory could be generalized, it would not be final. In short, if the strand model is a final description, the efforts of mathematical physicists just described must all be impossible.

Where does this fondness for generalization come from? In history of physics, generalizations often led to advances and discoveries. In the past, generalizations often led to descriptions that had a *wider range* of validity. As a result, generalizations became the way to search for new discoveries. Indeed, in the history of physics, the old theory often was a *special case* of the new theory. This relation was so common that usually, *approximation* and *special case* were taken to be synonyms.

General relativity and the standard model must be *approximations* of the final theory. But can either general relativity or the standard model of particle physics be *special cases* of the final, unified theory? Or, equivalently: Can the unified theory be a generalization of existing theories? The answer is no. The two existing theories cannot be *special cases* of the final theory. If the unified theory were a generalization of the two existing theories, it could not explain any of the millennium issues of Table 9! Indeed, if general relativity or the standard model of particle physics cannot explain the millennium issues, any generalization of them cannot either. Generalizations have no explanatory power. In other words, if the strand model is a unified description, it must not allow to deduce special cases. If the strand model is a final description, approximations of the strand model must exist, but special cases must not.

In summary, the final, unified description of motion must neither allow generalization nor must it itself be a generalization of either the standard model or of general relativity. The unified theory cannot be generalized and cannot be specialized; the unified theory must be *unmodifiable*.[*] This requirement is extremely strong; you may check that it eliminates most past attempts at unification. For example, this requirement suggests to eliminate grand unification, supersymmetry and higher dimensions as aspects of the final theory: indeed, these ideas are modifiable and they generalize the standard model of elementary particles; thus, these ideas lack explanatory power. *In short, a theory that is not final cannot be unified.* We will discover that the strand model is indeed *unmodifiable* and that it *explains* the standard model of elementary particles; so far, no modification of the strand model has been found. The strand model is indeed a candidate for the final theory.

Challenge 114 e

Challenge 115 e

WHY STRANDS? – SIMPLICITY

> Simplex sigillum veri.[**]
>
> Antiquity

Let us assume that it is not clear yet wether the strand model can be modified or not. Then there are still two reasons to explore featureless strands as basis for a unified description. First, featureless strands are the *simplest* known model that unifies quantum field theory and general relativity. Second, featureless strands are the *only known* model that realizes an important requirement: a unified description must not be based on points,

Page 100

Ref. 157

[*] Later I discovered that David Deutsch makes a similar point with his criterion that an explanation is only correct if it is *hard to vary*. Used in the case of a final theory, we can say that the final theory must be an *explanation* of general relativity and of the standard model. This implies that the final theory must be hard to vary. This matches the above conclusion that the final theory must be unmodifiable.

[**] ‘Simplicity is the seal of truth.’

sets or any axiomatic system. Let us explore the issue of simplicity first.

In order to reproduce three-dimensional space, Planck units, spin, and black-hole entropy, the fundamental constituents must be *extended* and *fluctuating*. We have deduced this result in detail in the previous chapter. The extension must be one-dimensional. This is the simplest option, and it is also the only option compatible with three-dimensional space. In fact, one-dimensional strands explain the three-dimensionality of space, because tangles of one-dimensional strands exist *only* in three spatial dimensions. In four or more dimensions, any tangle or knot can be undone; this does not occur in three spatial dimensions. Page 119

No *simpler* model than featureless strands is possible. Other extended constituents that have been explored – ribbons, bands, strings, membranes, posets, branched lines, networks, crystals, or quantum knots – all increase the complexity. In fact they do so in *two* ways: they increase the number of features of the fundamental constituents and they complicate the mapping from the model to observation. Ref. 143 Ref. 158 Ref. 159 Ref. 160

First, none of these other models uses *featureless* constituents. In these other models, the fundamental constituents have properties such as width, twists, orientation, field values, coordinates, quantum numbers, tension, non-trivial topological information, etc. In some models, space-time is non-commutative or fermionic. All these features are *assumed*; they are added to the model by fiat. As such, they are difficult if not impossible to justify. In addition, these features increase the complexity of the models, of what happens, and how it happens. In contrast, the strand model has no justification issue and no complexity issue.

Secondly, the link between these more involved models and experiment is often *intricate* and sometimes not unique. The difficulties resulting from superstrings are well-known. In contrast, the strand model argues that the experimentally accessible Dirac equation of quantum field theory and the experimentally accessible field equations of general relativity arise *directly*, from an averaging procedure of crossing switches. Indeed, the strand model proposes to unify these two halves of physics with only one fundamental principle: strand crossing switches define Planck units. In fact, we will find out that the strand model describes not only vacuum and matter, but also gauge interactions and particle properties as *natural* consequences of the microscopic structure of nature at Planck scales. The comparable ideas in other models are much more elaborate. Ref. 155

We remark that building three-dimensional physical space from strands is even simpler than building it from points! In order to build three-dimensional space from *points*, we need concepts such as sets, neighbourhood, topological structures, and metric structures. And despite all these intricate concepts, the concept of space defined in this way still has no defined physical length scale; in short, it is not the same as physical space. In contrast, in order to build three-dimensional physical space from *strands*, we need no fundamental points, sets, or metric structures; we only need long-time averages of strands and their crossings. And the length scale is built in.

All this suggests that the strand model, based on featureless, one-dimensional and fluctuating constituents, might be the model for unification with the smallest number of concepts, thus satisfying Occam's razor. In short, the strand model seems to be the *simplest* possible unified description. In fact, we will discover that strands are the simplest way to model particles, interactions and the vacuum, while fulfilling the requirements of a final theory.

TABLE 10 The differences between nature and any description.

NATURE	DESCRIPTION
Nature is not a set.	Descriptions needs sets to allow talking and thinking.
Nature has no events, no points and no continuity.	Descriptions need events, points and continuous $3 + 1$-dimensional space-time to allow formulating them.
Nature has no point particles.	Descriptions need point particles to allow talking and thinking.
Nature is not local.	Descriptions need locality to allow talking and thinking.
Nature has no background.	Descriptions need a background to allow talking and thinking.
Nature shows something akin to $R \leftrightarrow 1/R$ duality.	Descriptions need to break duality to allow talking and thinking.
Nature is not axiomatic.	Axiomatic descriptions are needed for precise talking and thinking.

The simplicity of a model helps in two ways. First, the simpler a model is, the freer it is of ideology, preconceptions and beliefs. Secondly, the simpler a model is, the easier it is to check it against observation. In particular, it is simple to check its solution of paradoxes.

WHY STRANDS? – THE FUNDAMENTAL CIRCULARITY OF PHYSICS

The strand model describes strands as fluctuating in a background space-time of three plus one space and time dimensions. The background space-time is introduced by the observer. The background is thus different for every observer; however, all such backgrounds have three dimensions of space and one of time. The observer – be it a machine, an animal or a human – is itself made of strands, so that in fact, the background space is itself the product of strands.

We therefore have a fundamental circular definition: we describe strands with a background, and the background with strands. *Strands thus do not provide an axiomatic system in the mathematical sense.* This fulfils one of the requirements for the unified description. Page 103

Why does the fundamental circular definition arise? Physics is talking (and thinking) about nature and motion. A unified model of physics is talking about motion with highest precision. This implies that on one hand, as talkers, we must use concepts that allow us *to talk*. Talking and thinking requires that we use continuous space and time: we must use a *background*. On the other hand, to talk *with precision*, we must have a minimum length, and use *strands*. There is no way to get rid of this double and apparently contradictory requirement. More such contradictory requirements are given in Table 10. And since there is no way to get rid of these contradictory requirements, we don't: *we use both*

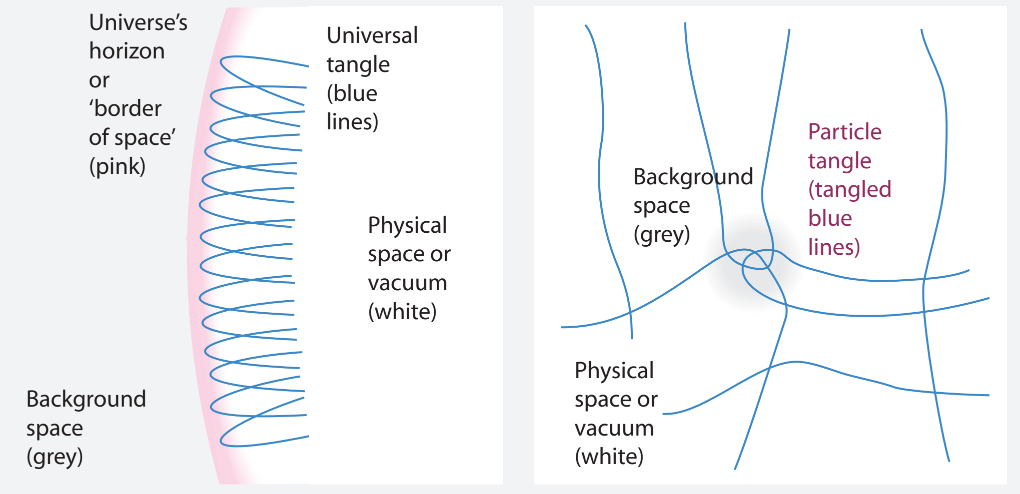


FIGURE 17 In the strand model, physical space – or vacuum – and background space are distinct, both near the horizon and near particles.

continuous background space-time and strands to describe nature.

A unified model of physics is talking about motion with highest precision. This forces us to use continuous space-time and strands at the same time. This is not a contradiction; it is (only) a circular definition. Since we, the talkers, are part of nature, a unified model means that we, the talkers, talk about ourselves.*

We do not state that background space and time exist *a priori*, as Immanuel Kant states, but only that background space and time are *necessary* for thinking and talking. Ref. 161 In fact, *physical* continuous space and time result from strands, and thus do not exist a priori; however, background space and time are required concepts for any description of observations, and thus necessary for thinking and talking.

We have always to be careful to keep the fundamental circular definition of strands and backgrounds *in our mind*. Any temptation to resolve it leads astray. For example, if we attempt to define *sets* or *elements* (or points) or with the help of measurements, we are hiding or forgetting the fundamental circularity. Indeed, many physicists constructed and still construct axiomatic systems for their field. The fundamental circularity implies that axiomatic systems are possible for *parts* of physics, but not for physics as a whole. Indeed, there are axiomatic descriptions of classical mechanics, of electrodynamics, of quantum theory, of quantum field theory, and even of general relativity. But there is no axiomatic system for *all* of physics.

Strands fluctuate in a background space. Only crossing switches can be observed. In particular, this means that the mathematical points of the background space cannot be observed. In other words, despite using mathematical points to describe the background space (and strands themselves), none of them have physical significance. *Physical points do not exist in the strand model.* Physical locations of events are due to crossing switches,

---

* It is essential that despite this circularity, Gödel's incompleteness theorem does not apply to the situation, as it does not apply to any unified theory of physics. The incompleteness theorem is based on self-referential statements. Such statements neither appear in physics, nor in the strand model, nor in most of mathematics.

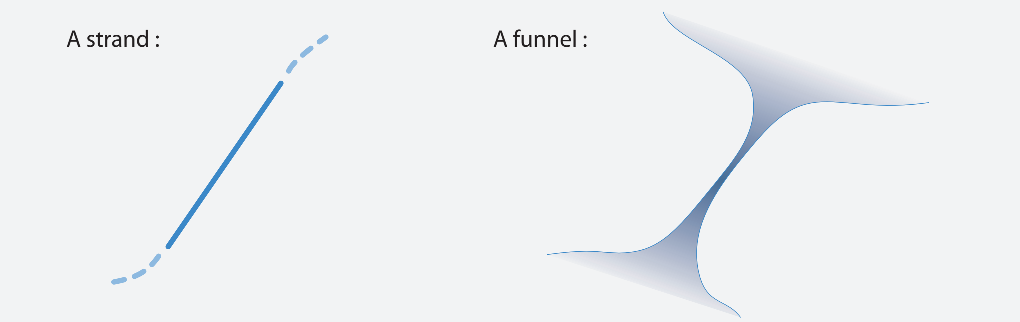


FIGURE 18 Two equivalent depictions of the fundamental constituents of nature: strands and funnels.

and can at best be localized to within a Planck length. The same limitation applies to physical events and to physical locations in time. A natural Planck-scale non-locality is built into the model. This realizes another requirement that any unified description has to fulfil.

The situation for physicists working on unification is thus harder – and more fascinating – than that for biologists, for example. Biology is talking about living systems. Biologists are themselves living systems. But in the case of biologists, this does *not* lead to a circular definition. Biology does not use concepts that contain circular definitions: a living being has no problems describing *other* living beings. Even neurobiologists, who aim to explore the functioning of the brain, face no fundamental limit doing so, even though they explore the human brain using their own brain: a brain has no problem describing other brains. In contrast, physicists working on unification need to live with Ref. 162 circularity: a fundamental, precise description of motion requires to be conscious of our own limitations as describing beings. And our main limitation is that we cannot think without continuous space and time, even though these concepts do not apply to nature.

We conclude: *A unified description cannot be axiomatic, cannot be based on observable physical points, must distinguish physical space from background space, and cannot be background-independent.* Most models based on extended constituents also use backgrounds. However, most models also allow the definition of sets and axiomatic descriptions. Such models thus cannot be candidates for a unified description of nature. In contrast, the strand model keeps the fundamental circularity of physics intact; it does not allow an axiomatic formulation of fundamental physics, nor the observation of points or sets.

An equivalent alternative to strands

Another type of constituent also fulfils all the conditions for a unified description. As shown in Figure 18, as an alternative to fluctuating strands, we can use fluctuating *funnels* as fundamental constituents.

Funnels are similar to wormholes; however, both their ends lead into usual three-dimensional space. Funnels are also similar to D-branes, except that they are embedded in three spatial dimensions. A funnel also resembles a projected part of an exotic manifold. In the funnel image, physical space is the superposition of all funnel openings.

Page 132 Fluctuating funnels also remind us of amoebas.

A quick check shows that the funnel alternative is *completely equivalent* to strands. You might enjoy testing that all the conclusions deduced in the following pages appear Challenge 116 e unchanged if strands are replaced by funnels. In particular, also funnels allow to deduce quantum field theory, the standard model and general relativity. However, the mentioned similarities of funnels with wormholes, D-branes or exotic manifolds are of no help: none of these approaches had led to viable models of unification. Due to the strict equivalence between strands and funnels, the choice between the two alternatives is a matter of taste or of visualization, but not a matter of physics. We use strands in the following, as they are simpler to draw.

Summary on the fundamental principle of the strand model – and on continuity

We have introduced featureless, fluctuating *strands* as common constituents of space, matter and radiation. All physical *observables* are defined through Planck units as consequences of crossing switches of strands. *Continuity* of any kind, of space, fields or time, results from the averaging of strand crossing switches.

The strand model fulfils the general requirements for a final and unified description listed in Table 7, provided that it describes all motion with precision.

At this point we must start the comparison with experiment. Even though we need to check that strands describe *all* motion, the task is limited: we only need to check whether strands solve each of the millennium issues given in Table 9. If the strand model can solve them, then it reproduces all observations and provides a unified description of nature. If not, the strand model is worthless.

Chapter 9

QUANTUM THEORY OF MATTER DEDUCED FROM STRANDS

We show in this chapter that featureless strands that fluctuate, together with the fundamental principle – the definition of $\hbar/2$ as crossing switch – imply without alternative that matter is described by quantum theory. We deduce that tangles of fluctuating strands reproduce the spin 1/2 behaviour of matter and allow to define wave functions. In particular, fluctuating strands imply that motion of matter follows the Dirac equation. Furthermore, we show that strands imply the least action principle, and therefore, that tangles of fluctuating strands are described by the Lagrangian of relativistic quantum particles.

In the present chapter, we derive the quantum theory of *matter*; thus we show that strands reproduce all observations about matter. We leave for later the derivation of quantum theory of light and radiation, the standard model of elementary particles, and the quantum description of gravitation. As usual in quantum theory, we work in *flat* space-time only.

Strands, vacuum and particles

In nature, particles move in the vacuum. Vacuum is free of matter and energy. In the strand model,

▷ ***Vacuum*** is a collection of fluctuating, unknotted and untangled strands.

The time average of such simple strands has no energy and no matter content, because there are (on average) no crossing switches. This is illustrated in Figure 19. The temporary switches that can appear through fluctuations of the vacuum turn out to be virtual particles; we will explore them below. We note that the (flat) vacuum, being a time average, is *continuous* and *unique*. The strand model thus contains both a minimum length and a continuous vacuum. The two aspects do not contradict each other.

In nature, particles *move*: particles change position and phase over time. We therefore must define these concepts.

▷ A ***particle*** is a tangle of strands.

▷ The ***position*** of a particle is given by the centre of the tangle core.

▷ The ***phase*** of a matter particle is given by half the angle that describes the orientation of the tangle core around the spin axis.

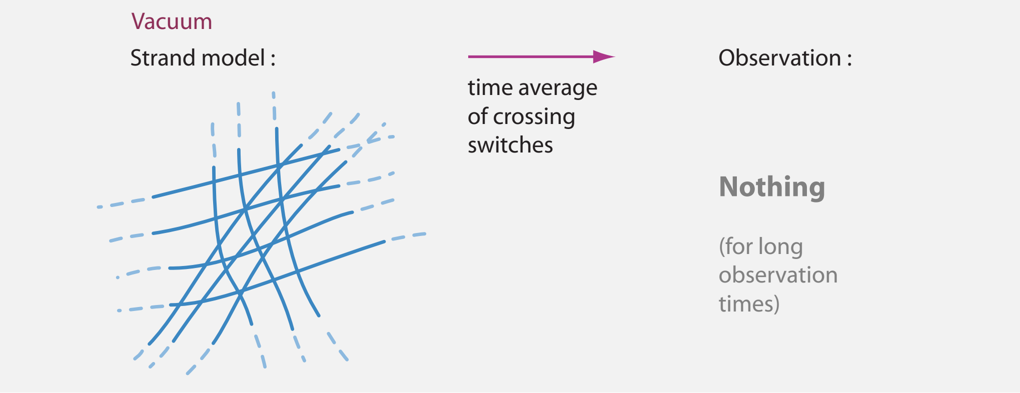


FIGURE 19 Vacuum as a collection of unknotted and untangled strands.

Page 173

- ▷ The ***wave function*** of a matter particle is a blurred rendering of its fluctuating strand crossings.

Page 173

These definitions are illustrated in Figure 20. We note that these three definitions imply a short-time average over tangle fluctuations that will be specified in detail below. The *tangle core* is the knotted part of the tangle, which contains all the links. The core is connected to the border of space by the *tails* of the tangle. We thus get:

- ▷ ***Motion*** of a particle is the change of the position and orientation of its tangle core.

In nature, quantum particle motion is described by *quantum theory*. The main property of quantum theory is that the invariant quantum of action $\hbar$ appears. In the strand model, $\hbar/2$ is described by a single crossing switch; the value of the quantum of action is thus invariant by definition.

We now explore how precisely the quantum of action enters in the motion of quantum particles. In particular, we will show that tangle fluctuations reproduce quantum theory. As an advance summary, we clarify that

- ▷ ***Quantum motion*** is due to fluctuations of tangle *tails*. The deformations of the tangle core are not important for quantum motion; for simplicity it is assumed that the core remains unchanged in quantum motion.

Page 207

We will study the deformations of cores in the next chapter, where we show that they are related to *interactions*. In this chapter we explore the deformations of tangle tails; they produce the motion of quantum particles – more precisely, the motion of *free* and *stable* quantum particles. We first study the rotation and then the translation of matter quantum particles.

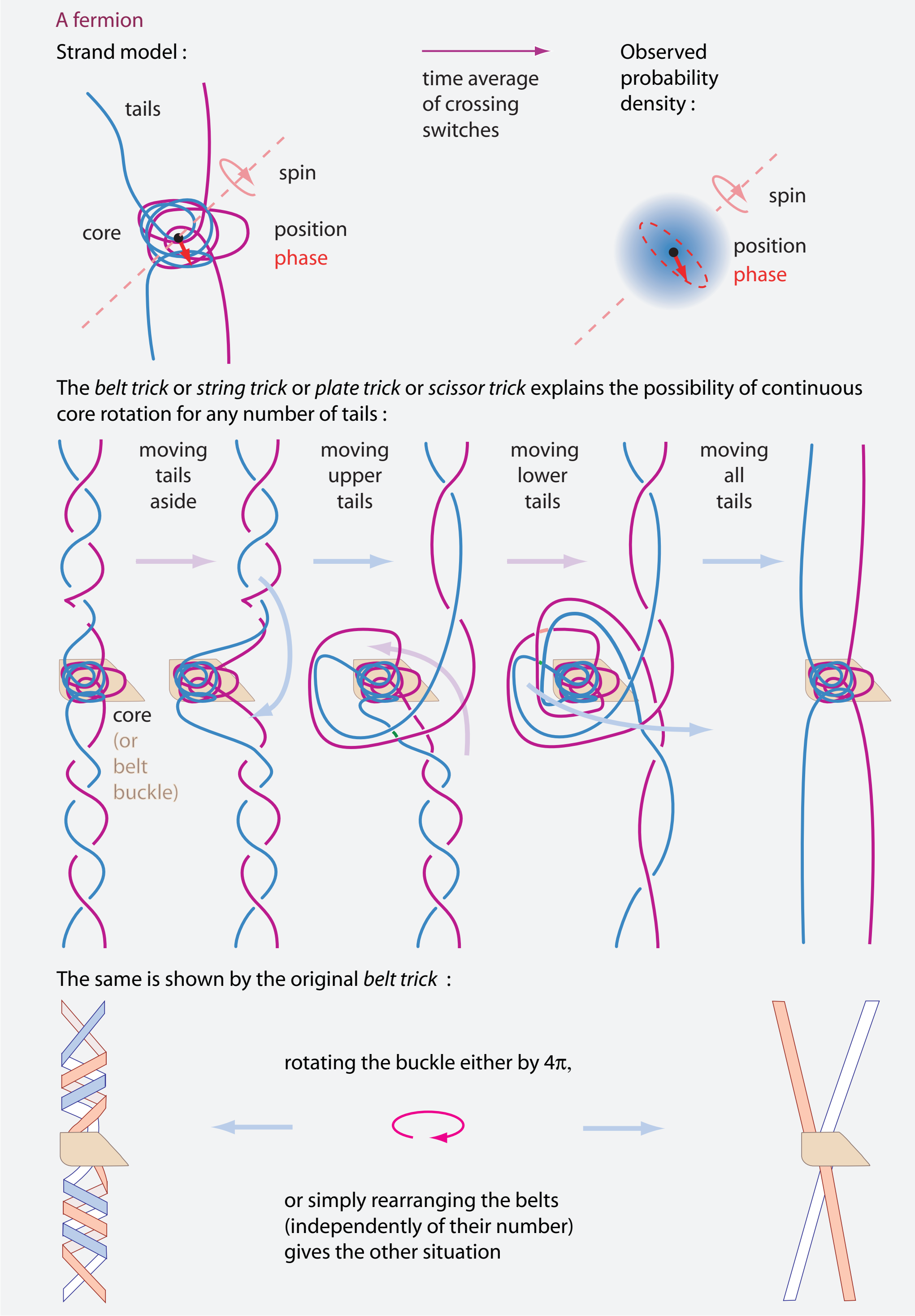


FIGURE 20 A spin 1/2 tangle with its core, tails, and the corresponding particle position and phase. Below, the *belt trick* or *scissor trick* or *string trick* or *plate trick*, shows that rotations by 4 π of an object with three or more tails (or with one or more ribbons) are equivalent to no rotation at all. This allows a suspended object, such as a belt buckle or a tangle core, to rotate for ever.

THE BELT TRICK, ROTATION AND SPIN 1/2

In nature, quantum particles have specific spin values and specific exchange properties. In particular, quantum particles have a spin axis, have a phase, and follow the spin–statistics theorem. We now show that all these properties follow from the strand model.

In nature, quantum particles are described by their behaviour under *rotation* and by their behaviour under *exchange*. The behaviour of a particle under rotations is described by its spin value. The behaviour under exchange can be of two types: a quantum particle can be a fermion or a boson. In nature, *particles with integer spin are bosons, and particles with half-integer spin are fermions*. This is the *spin–statistics theorem*. We now show that the spin–statistics theorem follows naturally from the strand model. We start with the case of spin 1/2 particles, and first clarify the nature of particle rotation. (We follow the usual convention to use 'spin 1/2' as a shorthand for 'z-component of spin with value $\hbar/2$'.)

It is often said that spin is *not* due to rotation. This misleading statement is due to two arguments that are repeated so often that they are rarely questioned. First, spin 1/2 particles cannot be modelled as small rotating stones. Secondly, it is impossible to imagine rotating electric charge distributions with a speed of rotation below that of light and an electrostatic energy below the observed particle masses. Both statements are correct. Despite being correct, there is a way to get around them; at the present stage, we focus on the first: we will show that spin *can* be modelled as rotation.

In the strand model, for all quantum particles we have:

▷ ***Spin*** is core rotation.

Indeed, in the strand model, all quantum particles, including those with spin 1/2, *differ* from everyday objects such as stones, and the essential difference is due to extension:

▷ ***Quantum particles*** are particles whose tails cannot be neglected.

In everyday objects such as stones, even though they are composed of quantum particles, tails do not play an important role, because everyday objects are mixed states, and not eigenstates of angular momentum. In everyday objects, tails can be neglected. But for quantum particles, the tails are essential. We will see step by step that the tails of quantum particles explain their spin behaviour, their exchange behaviour, and their wave behaviour. We will see below that in the strand model, wave functions are *blurred* tangles; we can thus explore the general behaviour of wave functions by exploring the behaviour of tangles.

It has been known for many decades that so-called *belt trick*, illustrated in Figure 20, Figure 21 and Figure 22, can be used, together with its variations, to model the behaviour of spin 1/2 particles under rotations. The belt trick is the observation that a belt buckle rotated by *two* full turns – in contrast to a buckle rotated by only *one* full turn – can be brought back into its original state without moving the buckle; only the motion of the belt is necessary. The belt trick is also called the *scissor trick*, the *plate trick*, the *string trick*, the *Philippine wine dance* or the *Balinese candle dance*. It is sometimes incorrectly attributed to Dirac. Ref. 163

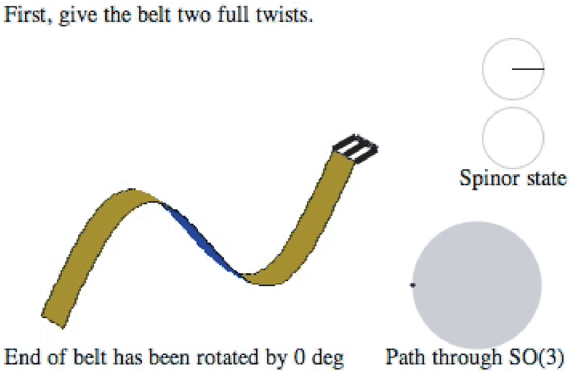


FIGURE 21 The belt trick: a double rotation of the belt buckle is equivalent to no rotation; this animation shows the *first* way in which the belt trick can be performed. Not shown: the belt trick is also possible with *any* number of belts attached to the buckle. (QuickTime film © Greg Egan)

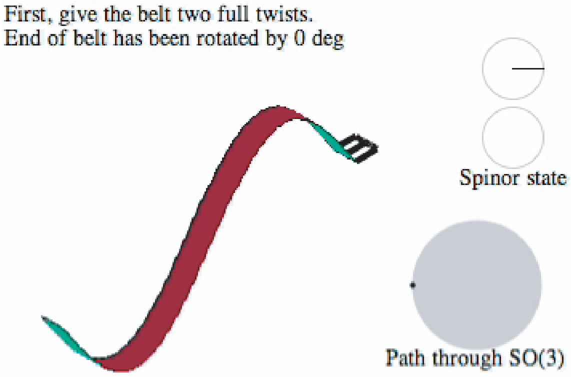


FIGURE 22 The belt trick again: this animation shows the *second* way in which the belt trick can be performed. Not shown: the belt trick is also possible with *any* number of belts attached to the buckle. (QuickTime film © Greg Egan)

The belt trick is of central importance in the strand model of spin 1/2 particles. In the strand model, all spin 1/2 particles are made of *two* (or more) tangled strands, and thus have four (or more) tails to the 'border', as shown in Figure 20. For such tangles, a rotation by 4π of the tangle core – thus a rotation by *two* full turns – can bring back the tangle to the original state, provided that the tails can fluctuate. Any system that returns to its original state after rotation by 4π is described by spin 1/2. (In fact, the tails must be unobservable for this equivalence to hold; in the strand model, tails are single strands and thus are unobservable.) We will show below that the intermediate twisting of the tails that appears after rotation by 2π corresponds to a multiplication of the wave function by −1, again as expected from a spin 1/2 particle. Page 179

The belt trick thus allows a pointed object or a tangle core that is attached by (three or more) strands to the border of space to rotate *continuously*. The possibility of continuous rotation allows us to describe spin 1/2 particles by rotating tangles. In other terms, rotating tangles model spin. The fluctuations required to rearrange the tails after two full turns can be seen to model the average *precession* of the spin axis. We thus confirm that spin and rotation are the same for spin 1/2 particles.

Ref. 163 We stress an aspect of the belt trick that seems unmentioned in the literature: after a

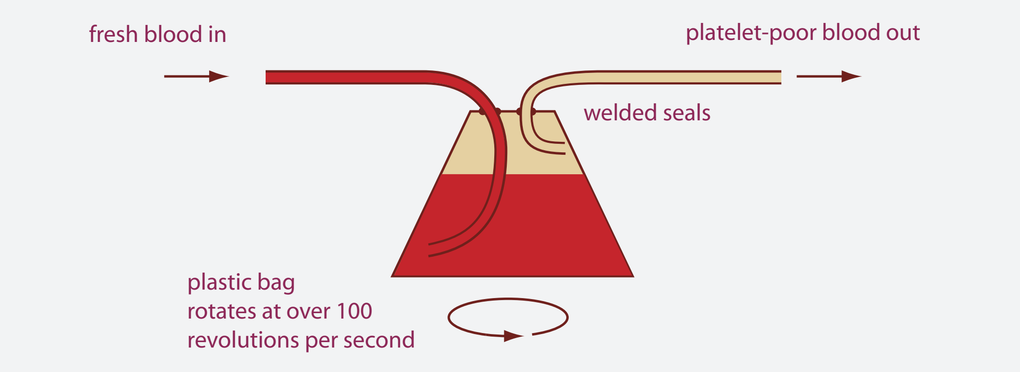


FIGURE 23 In an apheresis machine, the central bag spins at high speed despite being connected with tubes to a patient.

rotation of the belt buckle or tangle core by 4π, there are *two* options to untangle the tails. The two options are shown in Figure 21 and Figure 22. You can test this yourself, using a real belt. In short, there are *two* ways to perform the belt trick. This two-fold option will be of central importance later on: the two options distinguish matter from antimatter.

Challenge 117 e

An aside: the belt trick saves lives

Without the belt trick, the *apheresis machines* found in many hospitals would not work. When a person donates blood platelets, the blood is continuously extracted from one arm and fed into a bag in a centrifuge, where the platelets are retained. The platelet-free blood flows back into the other arm. This happens continuously, for about an hour or two. In order to be sterile, tubes and bag are effectively one piece and used only once, as shown in Figure 23. Topologically, this set-up is identical to a central tangle with at least 2 strands, i.e., at least 4 tails, thus to the tangle model of an elementary fermion.

In such apheresis machines, centrifugation takes place at over 100 revolutions per second. To avoid tangling up the blood tubes, a bracket moves the tubes during each rotation, alternatively up and down, in the same way that the belt moves when the buckle is rotated by 4π. The platelets are retained in the bag. They are then used to treat patients with leukaemia or severe blood loss due to injury. A single platelet donation can sustain several lives. In short, the belt trick allows sterile, and thus safe, platelet donations that save other people's lives.

Fermions, spin and statistics

Fermions are particles whose wave function changes sign when they are exchanged. As mentioned, we will see below that in the strand model, wave functions are *blurred* tangles. We thus can explore exchange properties of quantum particles by exploring the exchange properties of their tangles.

The exchange properties of spin 1/2 tangles are easily checked by playing around with some pieces of rope or bands, as shown in Figure 24. If we exchange two tangle cores *twice*, while keeping all tails connections fixed, tail fluctuations alone can return the situation back to the original state.

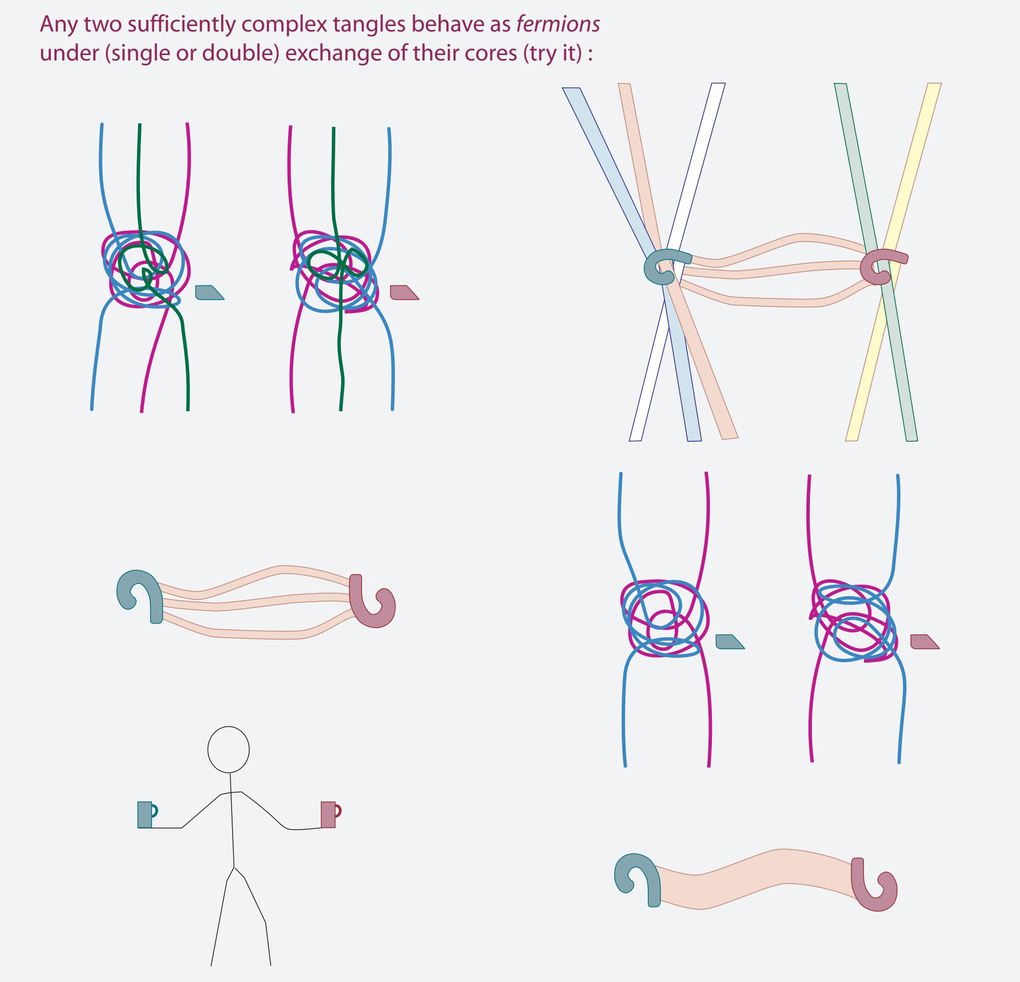


FIGURE 24 When two spin 1/2 tangles, each made of several strands or bands, are exchanged *twice*, it is possible to rearrange their tails to yield the original situation. This is not possible when the tangles are only rearranged *once*. Spin 1/2 tangles are thus fermions.

The simplest possible version of this experiment is the following: take two coffee cups, one in each hand, and cross the two arms over each other (once). Keeping the orientation of the cups fixed in space, uncross the arms by walking around the cups. This is possible, but as a result, both arms are twisted. If you are intrepid, you can repeat this with two (or more) people holding the cups. And you can check the difference with what is possible after a double crossing of arms.

These experiments show that a *simple* exchange of two spin 1/2 particles (tangles, cups on hands, belt buckles) is equivalent to a multiplication by -1 (twisted tangles, arms or belts). In contrast, a *double* exchange of two spin 1/2 particles can always be untwisted and is equivalent to no exchange at all. Spin 1/2 particles are thus fermions.

In summary, a tangle core made of two or more tangled strands behaves both under rotations and under exchange like a spin 1/2 particle. The strand model reproduces the spin–statistics theorem for spin 1/2.

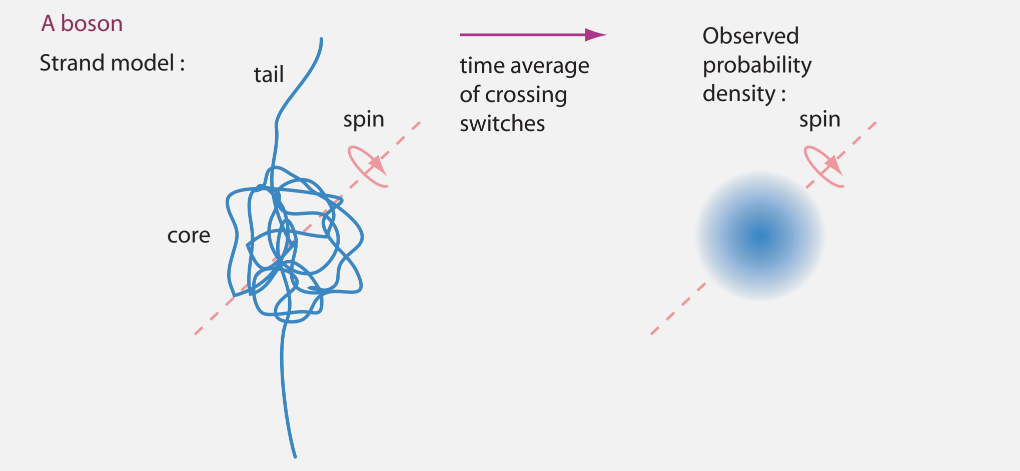


FIGURE 25 A massive spin 1 *boson* in the strand model (left) and the observed probability density when averaging its crossings over long time scales (right).

Bosons, spin and statistics

For tangles made of *one* strand – thus with two tails to the border – a rotation by 2π restores the original state. Such a tangle, shown in Figure 25, behaves like a spin 1 particle, thus like a boson – also under exchange.

A strand model for the graviton, a boson invariant under rotations by π and thus with spin 2, will be introduced in the chapter on general relativity. Page 248

In short, the spin–statistics theorem for all elementary particles, fermions or bosons, can be reproduced by fluctuating strands, depending on the tangle details. The strand model also implies that no spins lower than $\hbar/2$ are possible, and that spin values are always an *integer multiples of* $\hbar/2$. Challenge 118 e

In the strand model, temporal evolution and particle reactions *conserve* spin, because all interactions conserve the number of strands and tails, as will become clear later on.

In passing, we have thus also explained the origin of permutation symmetry in nature. *Permutation symmetry* of particles is due the possibility to exchange tangle cores; identical particles have tangle cores of identical topology. We have thus already ticked off one item from the millennium list of unexplained properties of nature. Page 158

In summary, the strand model reproduces the rotation, the spin and the exchange behaviour of quantum particles, both fermions and bosons, in all its observed details. We now proceed to study the translational motion of quantum particles.

Tangle functions: blurred tangles

Strands and tangles are not observable. Only crossing switches are. To study crossing switches, we first recall what a crossing is.

▷ A ***crossing*** of strands is a local minimum of strand distance. The position, orientation and the phase of a crossing are defined by the space vector corresponding to the local minimum, as shown in Figure 26. The sign of the ori-

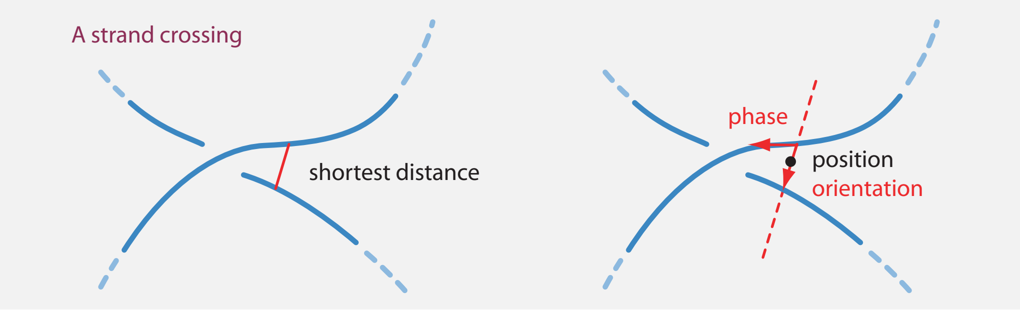


FIGURE 26 The definition of a crossing, its position, its orientation, and its phase.

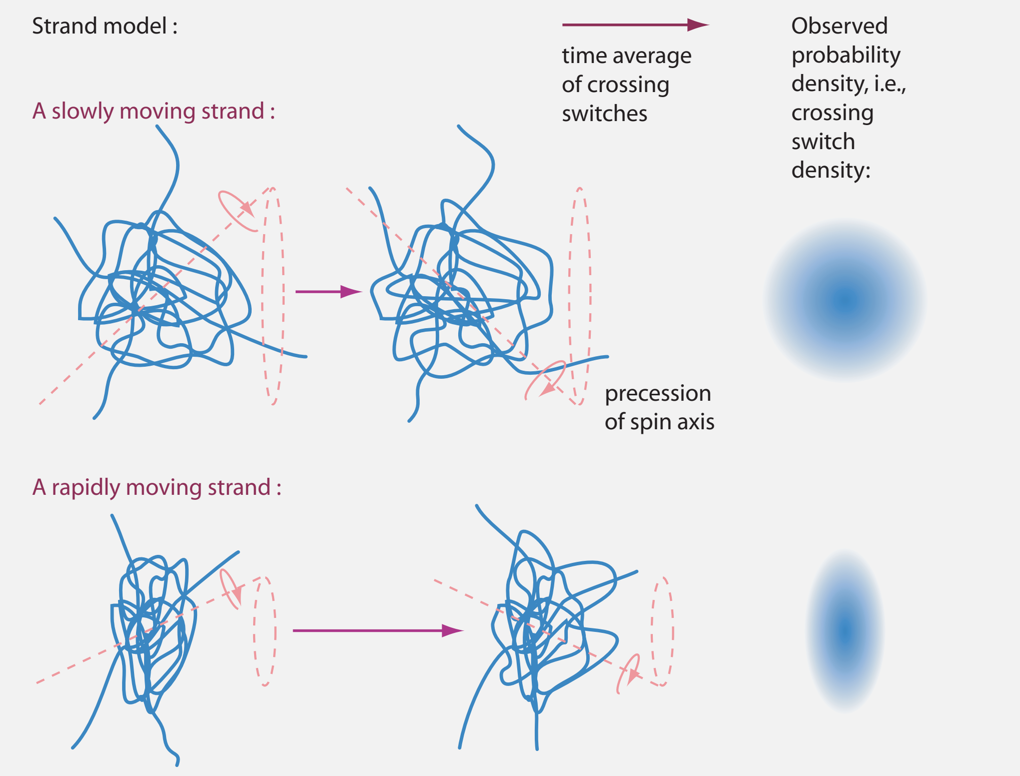


FIGURE 27 Some strand configurations, some of their short time fluctuations, and the corresponding probability density that results when averaging crossing switches over time.

entation is defined by arbitrarily selecting one strand as the starting strand. The arbitraryness in the definition of the phase will be of great importance later on.

To describe the motion of tangles, we need concepts that allow us to take the step from general strand fluctuations to the motion of tangle cores. As a mathematical tool to describe crossing fluctuations, we define:

▷ The ***tangle function*** of a system described by a tangle is the short-time average of the positions and the orientations of its crossings (and thus *not* of crossing switches).

The tangle function can be called the 'oriented crossing density' or simply the 'blurred tangle'. As such, the tangle function is a continuous function of space, similar to a cloud; we will see below what its precise mathematical description looks like. The tangle function captures the short-time average of all possible tangle fluctuations. For a tangle made of two strands, Figure 27 illustrates the idea. However, the right-hand side of the figure does not show the tangle function itself, but its probability density. We will see shortly that the probability density is the (square of the) crossing *position* density, whereas the tangle function is a density that describes *both position and orientation* of crossings.

The tangle function at any given time is *not* observable, as the definition is not based on crossing switches, but only on crossings. However, since crossing switches only occur at places with crossings, the tangle function is a useful tool to *calculate* observables. In fact, we will show that the tangle function is just another name for what is usually called the *wave function*. In short, the tangle function, i.e., the oriented crossing density, will turn out to describe the *quantum state* of a system.

In summary, the tangle function is a blurred image of the tangle – with the important detail that the crossings are blurred, not the strands.

▷ For the definition of the tangle function, the ***short-time average*** of crossings is taken over the typical time resolution of the observer. This is a time that is much *longer* than the Planck time, but also much *shorter* than the typical evolution time of the system. The time resolution is thus what the observer calls an 'instant' of time. Typically, this will be 10^{-25} s or more; the typical averaging will thus be over all times between 10^{-43} s, the Planck time, and 10^{-25} s or more.

There are *two* ways to imagine tangle fluctuations and to deduce the short-time average from a given tangle. The straight-forward way is to average over all possible strand fluctuations during the short time. *Each piece of strand* can change in shape, and as a result, we get a cloud. This is the common *Schrödinger view* of the wave function and of quantum mechanics. The alternative way to average is to imagine that the *tangle core as a whole* changes position and orientation randomly. This is easiest if the core with all its crossings is imagined to be tightened to a small, almost 'point-like' region. Then all observables are also localized in that region. It is often simpler to imagine an average over all position and orientation fluctuations of such a tightened core, that to imagine an average over all possible strand fluctuations. This alternate view leads to what physicists call the *path integral formulation* of quantum mechanics. (Can you show the equivalence of the two averaging methods?) Of course, the final result is again that the tangle function is a cloud. Challenge 119 ny

DETAILS ON FLUCTUATIONS AND AVERAGES

In the strand model, strand fluctuations randomly add detours to strands or randomly shift the core position. Fluctuations do not keep the strand length constant. Fluctuations do not conserve strand shape nor any other property of strands, as there is no mechanism that enforces such rules. Strand fluctuations are thus quite wild. What can be said about the details of the averaging procedure for strand fluctuations?

The *fluctuations of the vacuum* strands are those fluctuations that lead to the definition of the background space. This is possible in a consistent manner only if the fluctuations are homogeneous and isotropic. The vacuum state can thus be *defined* as that state for which the fluctuations are (locally) homogeneous and isotropic. The existence of a homogeneous and isotropic background space then implies conservation of energy, linear and angular momentum.

The *fluctuations of a tangle* lead, after averaging, to the tangle function, i.e., to the wave function. The conservation of energy and momentum implies that the time average of the tangle fluctuations also conserves these quantities.

Therefore we can continue our discussion without knowing the precise evolution of the tangle fluctuations themselves. We only assume that the average of the fluctuations behaves in such a way as to be *consistent* with the definition of the background used by the observer. We thus make explicit use of the conviction that a background-free description of nature is impossible, and that a fundamental description of nature *must* contain a circular definition that makes an axiomatic description of nature impossible.

Page 103, page 162

We will also show below that the definition of tangle function does *not* introduce hidden variables, even though first impression might suggest the opposite.

Page 193

TANGLE FUNCTIONS ARE WAVE FUNCTIONS

In the following, we show that the tangle function, the blurred image of tangle crossings, is the same as what is usually called the wave function. We recall what we know from textbook quantum theory:

▷ A single-particle wave function is, generally speaking, a *rotating and diffusing cloud*.

The *rotation* describes the evolution of the phase, and the *diffusion* describes the evolution of the density. We now show that tangle functions have these and all other known properties of wave functions. We proceed by deducing all the properties from the definition of tangle functions. We recall that, being a short-time average, a tangle function is a continuous function of space and time.

▷ Using the tangle function, we define the strand ***crossing position density***, or ***crossing density***, for each point in space, by discarding the orientation information. The crossing position density is a *positive number*, more precisely, a positive real function $R(x, t)$ of space and time.

We will see shortly that the crossing position density is the square root of what is usually called the *probability density*.

▷ A tangle function defines an ***average crossing orientation*** and a ***average phase*** at each point in space. The average crossing orientation and the average phase are related to the *spin orientation* and *phase* of the wave function. The mathematical descriptions of these quantities depend on the approximation used.

The *simplest approximation* for a tangle function is to assume, in the physical situation under study, that the spin direction is *independent* of spatial position; this approximation will lead to the Schrödinger equation. In this simplest approximation, at each point in space, the local average orientation of the fluctuations of the tangle core will just be described by a *single angle*. This quantum phase is a function of time and space and describes how much the local average phase is rotated around the fixed spin orientation.

▷ The ***quantum phase*** of fermions is *one half* the core rotation angle α.

The *more complicated cases*, when the spin axis can *change* over space, require more complicated descriptions of the orientation and phase averages; we will study these cases separately below. They will lead to the non-relativistic Pauli equation and to the relativistic Dirac equation.

In short, in the simple case when spin effects can be neglected, the local tangle function value can be described by one real number R and by one quantum phase α. The tangle function can thus be described by a *complex number* ψ at each point in space and time:

$$\psi(x,t) = R(x,t)\mathrm{e}^{i\alpha(x,t)/2} \; . \tag{134}$$

If a system changes with time, the tangle function changes; this leads to crossing switches; therefore, temporal evolution is expected to be observable through these crossing switches. As we will see shortly, this leads to an evolution equation for tangle functions.

Page 188 If *many* particles need to be described, the multi-particle tangle function defines a separate crossing density for each particle.

Tangle functions form a *vector space*. To show this, we need to define the ***linear combination*** or ***superposition*** $\chi = a_1\psi_1 + a_2\psi_2$ of two tangle functions. This requires the definition of two operations: scalar multiplication and addition. We can do this in two ways. The first way is to define the operations for tangle functions directly, as is done in quantum mechanics:

▷ First, boring definition: The ***scalar multiplication*** $a\psi$ and the ***addition*** $\psi_1 + \psi_2$ of quantum states are taken by applying the relative operations on complex numbers at each point in space, i.e., on the local values of the tangle function.

The second way to deduce the vector space is more fun, because it will help us to visualize quantum mechanics. We can also define all operations for tangles, and imagine the time average taken *after* the tangle operation is performed.

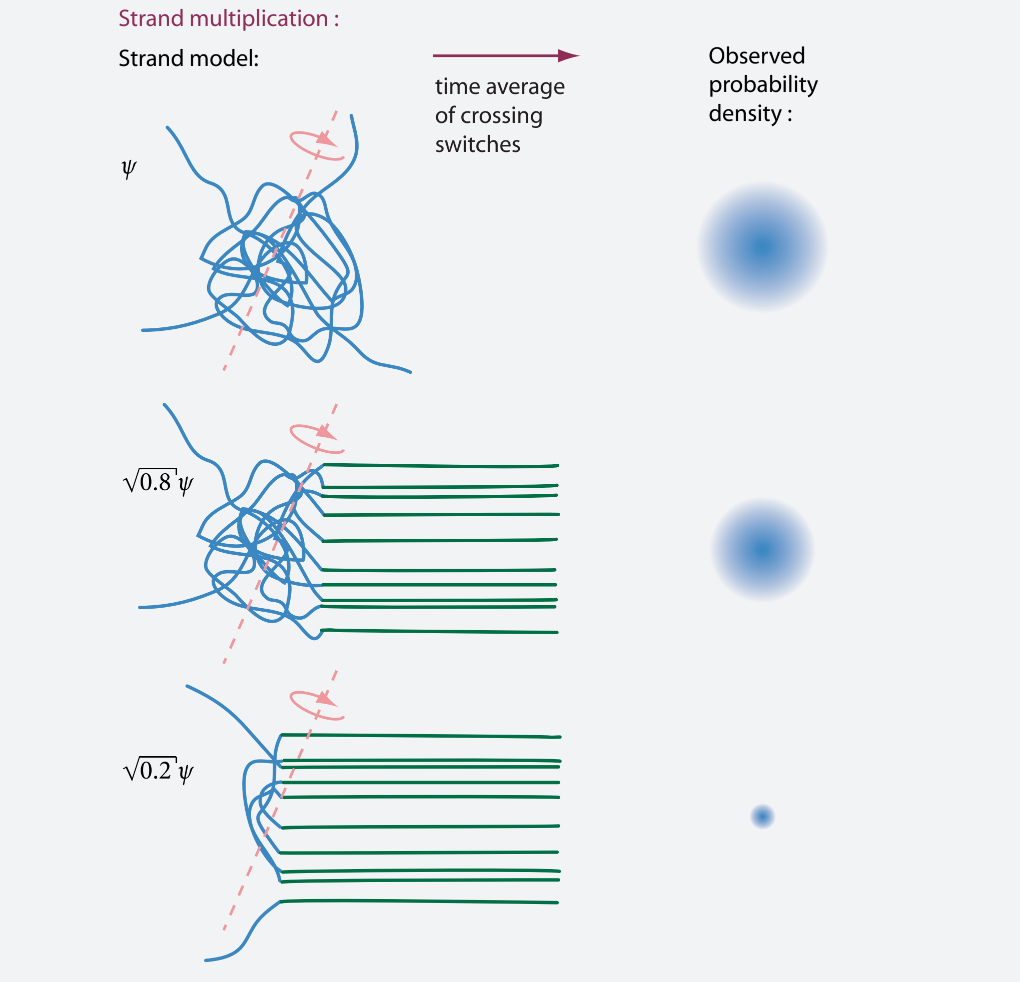


FIGURE 28 Scalar multiplication of localized tangles, visualizing the scalar multiplication of wave functions.

- ▷ Second, fun definition: The ***scalar multiplication*** $a\psi$ by a complex number $a = re^{i\delta}$ is formed by taking a tangle underlying the tangle function ψ, first rotating the tangle core by the angle 2δ, and then pushing a fraction $1 - r$ of the tangle to the border of space, thus keeping the fraction r of the original tangle at finite distances. Time averaging then leads to the tangle function $a\psi$.

The scalar multiplication for strands is illustrated in Figure 28. The above definition of scalar multiplication is defined for factors $r \leqslant 1$. Indeed, no other factors ever appear in physical problems (provided all wave functions are normalized), so that scalar multiplication is not required for other cases.

The strand version of scalar multiplication is *unique*; indeed, even though there is a choice about which fraction r of a tangle is kept and which fraction $1 - r$ is sent to the border of space, the resulting tangle function, which is defined as an average over fluctuations, is independent from this choice.

The scalar multiplication of strands behaves as expected for 1 and 0. By construction, the strand version of scalar multiplication is associative: we have $a(b\psi) = (ab)\psi$. The strand multiplication by -1 is defined as the rotation of the full tangle core by 2π.

We also need to define the addition operation that appears in the linear combination of two tangle functions. This is a straightforward complex addition at each point in space. Again, for fun, we also define the operation on tangles themselves, and take the time average that leads to the tangle function afterwards.

▷ Second, fun definition: The ***addition*** of two tangles $a_1\psi_1$ and $a_2\psi_2$, where ψ_1 and ψ_2 have the same topology and where $a_1^2 + a_2^2 = 1$, is defined by connecting those tails that reach the border of space, and discarding all parts of the tangles that were pushed to the border of space. The connection of tangles must be performed in such a way as to maintain the topology of the original tangles; in particular, the connection must not introduce any crossings or linking. Time averaging then leads to the tangle function of the superposition $\chi = a_1\psi_1 + a_2\psi_2$.

To visualize the result of addition and superposition, it is easiest to imagine that the strands reaching the border of space have fluctuated back to finite distances. This is possible because by definition, these connections are all unlinked. An example of superposition, for the case of two quantum states at different positions in space, is shown in Figure 29. We note that despite the wording of the definition, no strand is actually cut or reglued in the operation of addition.

Page 179

The definition requires that the strand χ that results from the linear combination has the same topology and the same norm as each of the two strands ψ_1 and ψ_2 to be added. Physically, this means that only states for the same particle can be added and that particle number is preserved; this implements the so-called *superselection rules* of quantum theory. The result is pretty, because in usual quantum mechanics the superselection rules need to be added by hand. This is not necessary in the strand model. It is also possible to extend the definitions of scalar multiplication and of addition to all complex numbers and to unnormed states, but this leads us too far from our story.

Challenge 120 e

The sum of two tangle functions is *unique*, for the same reasons given in the case of scalar multiplication. The definition of addition can also be extended to more than two terms. Addition is commutative and associative, and there is a zero state, or identity element, given by no strands at all. The definition of addition also implies distributivity with respect to addition of states and with respect to addition of scalars.

In short, tangle functions form a vector space. We now define the scalar product and the probability density in the same way as for wave functions.

▷ The ***scalar product*** between two states φ and ψ is $\langle\varphi|\psi\rangle = \int \overline{\varphi}(\mathbf{x})\psi(\mathbf{x})\mathrm{d}\mathbf{x}$.

▷ The ***norm*** of a state is $\|\psi\| = \sqrt{\langle\psi|\psi\rangle}$.

▷ The ***probability density*** ρ is $\rho(x,t) = \overline{\psi}(x,t)\psi(x,t) = R^2(x,t)$. It thus ignores the orientation of the crossings and is the square of the *crossing pos-*

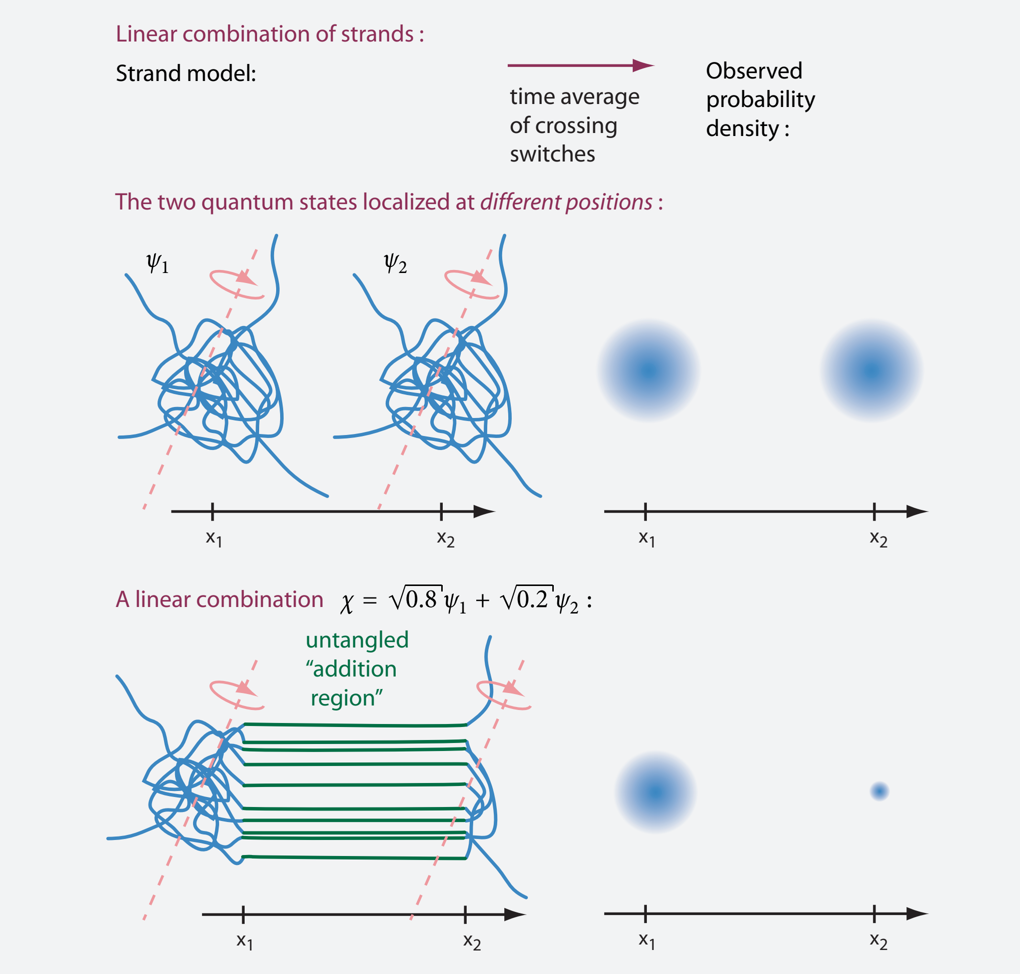


FIGURE 29 A linear combination of strands, in this case for two states representing a particle at two different position in space, visualizing the linear combination of wave functions.

ition density.

The scalar product and the probability density are *observables*, because their definitions can be interpreted in terms of crossing switches. Indeed, the scalar product $\langle\varphi|\psi\rangle$ can be seen as the (suitably normed) number of crossing switches required to transform the tangle $\overline{\varphi}$ into the tangle ψ, where the tangle $\overline{\varphi}$ is formed from the tangle φ by exchanging the orientation of each crossing. A similar interpretation is possible for the probability density, which therefore is at the same time the crossing density squared and the crossing switch density. We leave this confirmation as fun for the reader.

Challenge 121 e

It is also possible to define the scalar product, the norm and the probability density using tangles, instead of using tangle functions. This is left as a puzzle to the reader.

Challenge 122 ny

In summary, we have shown that tangle functions form a *Hilbert space*. The remaining steps are clear: We must first show that tangle functions obey the Schrödinger equation. Then we must extend the definition of quantum states by including spin and special rel-

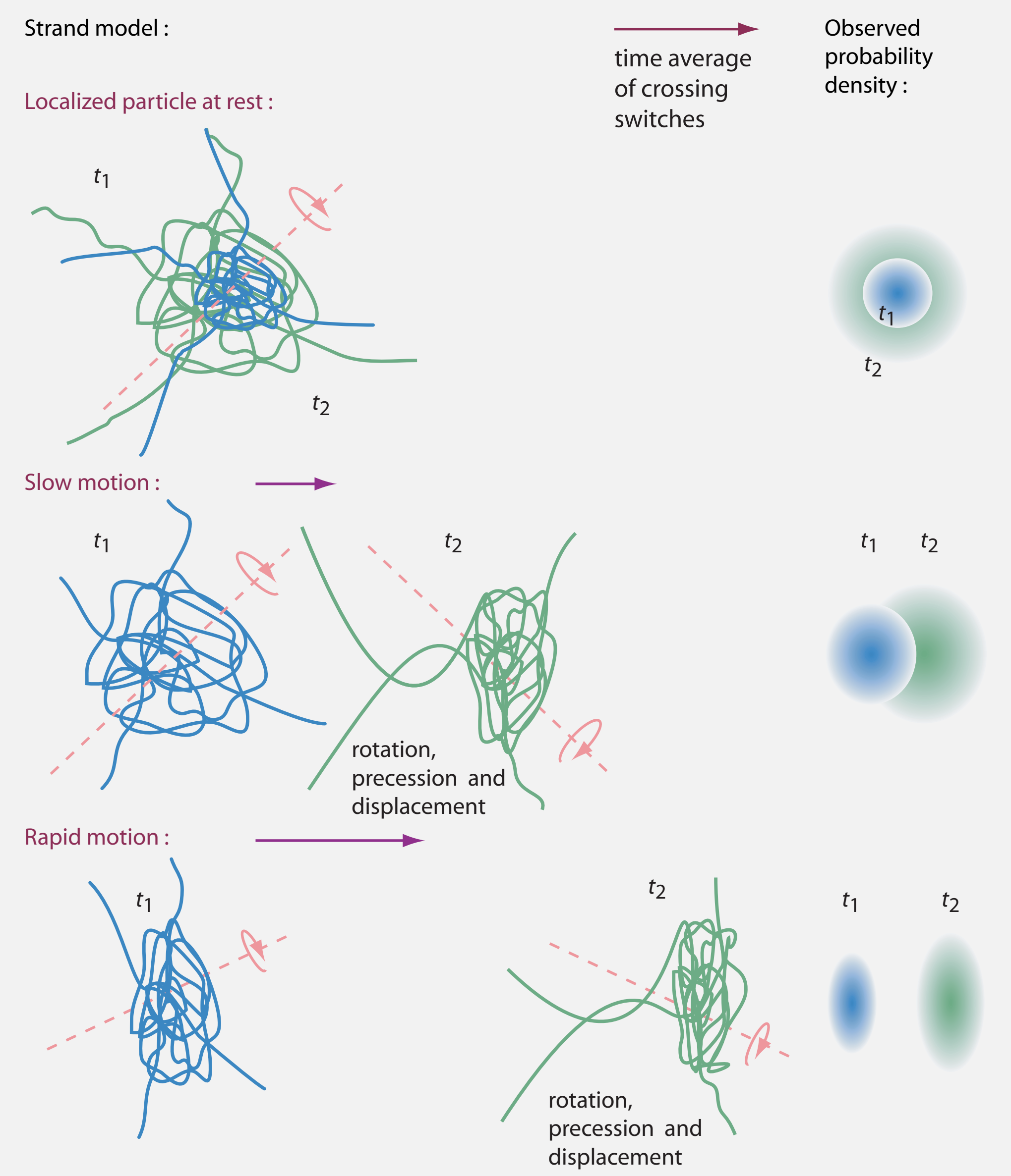


FIGURE 30 Examples of moving tangles of free particles.

ativity, and show that they obey the Dirac equation.

Deducing the Schrödinger equation from tangles

The Schrödinger equation, like all evolution equation in the quantum domain, results when the definition of the wave function is combined with the energy–momentum relation. As already mentioned, the Schrödinger equation for a quantum particle also assumes that the orientation of particle spin is constant for all positions and all times. In this case, the spin can be neglected, and the tangle function is a single complex number at each point in space and in time, usually written $\psi(x, t)$. How does the tangle function evolve in time? To answer the question, we only need the fundamental principle: crossing switches define the quantum of action $\hbar$.

We start with a free particle. We assume a fixed, but unspecified rotation direction of its tangle. In the strand model, a localized particle with constant speed is described by a localized tangle that rotates and advances. In other words, the strand fluctuations produce a peak of probability density that changes position with constant speed.

Every tangle rotation leads to crossing switches. A rapid tangle rotation leads to many crossing switches per time, and slow rotation to few crossing switches per time. Now, the fundamental principle tells us that crossing switches per time are naturally measured in action per time, or ***energy***. In other words, tangle rotation is related to tangle energy. Particles with high energy have rapidly rotating tangles, particles with low energy have slowly rotating tangles.

The energy of a rotating tangle is the number of crossing switches per time. The energy E is thus related to the angular frequency ω of the rotation by

$$E = \hbar\omega \ . \tag{135}$$

The local phase of the tangle function ψ changes with the rotation. This implies that

$$\omega = i\partial_t\psi \ . \tag{136}$$

We will need the relation shortly.

The linear motion of a tangle implies that it makes sense to speak of the number of crossing switches per distance. Rapidly moving tangles show many crossing switches per distance, slowly moving tangles show few crossing switches per distance. The fundamental principle tells us that the natural observable to measure crossing switches per distance is action per distance, or ***momentum***. Linear motion of tangles is related to momentum. The momentum of a moving tangle is the number of crossing switches per distance. The momentum p is thus related to the wave number $k = 2\pi/\lambda$ of the motion by

$$p = \hbar k \ . \tag{137}$$

The local phase of the tangle function ψ changes with the motion. This implies

$$k = -i\partial_x\psi \ . \tag{138}$$

We can now use the same argument that was used already by Schrödinger. The experimental dispersion relation for masses moving at velocities much smaller than the speed of light is

$$E = \frac{p^2}{2m} \ . \tag{139}$$

In fact, the strand model also allows to understand the dispersion relation itself. With increasing linear momentum, the spin rotation axis starts to align with the direction of motion. This leads to a quadratic increase of crossing switches with momentum p: one factor p is due to the increase of the speed of rotation, the other factor is due to the increase of the alignment. Substituting the above in the dispersion relation, we get the evolution equation for the tangle function ψ given by

$$i\hbar\partial_t\psi = -\frac{\hbar^2}{2m}\partial_{xx}\psi \, . \tag{140}$$

This is the famous Schrödinger equation for a free particle (written for just one space dimension for simplicity). We thus have deduced the equation from the strand model under the condition that spin can be neglected and that velocities are small compared to the speed of light.

At the same time, we have now completed the proof that tangle functions, in the case of negligible spin effects and small velocities, are indeed wave functions. In fact, tangle functions are wave functions also in the more general case, but then their mathematical description is more involved, as we will see shortly. In sloppy language, we have shown that *wave functions are blurred tangles.*

Mass from tangles

How fast does a tangle core rotate? This result of the belt trick will depend on the fluctuations of the core; these fluctuations determine an average angular frequency ω. What is the average size of the belt trick? The size will depend on the fluctuations of the core; these fluctuations determine the average wave number k. The proportionality factor $m = p^2/2E = \hbar k^2/2\omega$ is thus a quantity that relates rotation frequency and wave number, or the size and the speed of the belt trick. The quantity m, the mass, will depend on the fluctuation behaviour, which in turn will depend on the tangle topology. We note that a large mass value implies both slow rotation and small size.

In other words, the strand model links the ***mass*** m of a tangle to the tangle topology. The strand model thus promises to make particle masses *calculable*, as soon as the tangle topology is known. At this point, however, we are still in the dark. The missing steps are clear, however: first, we need to determine the tangle for each elementary particle, and then, we need to deduce particle mass values. This is our aim in the following.

Potentials

In quantum mechanics, interactions are described by potentials. An *electric potential* $V(x)$ changes the total energy of a particle with charge q at position x, since in quantum mechanics, electrostatic potentials influence the rotation velocity of the wave function. As a result, the left hand side of the Schrödinger equation (140), the energy term, is changed to $(\hbar\omega - qV)\psi(x, t)$.

Another possibility is a potential that does not change the rotation velocity, but that changes the wavelength of a charged particle. Such a *magnetic vector potential* thus changes the momentum term $\hbar\boldsymbol{k}$ on the right hand side of Schrödinger's equation to $(\hbar\boldsymbol{k} - q\boldsymbol{A})\psi(x, t)$. This double substitution, the so-called *minimal coupling*, is equivalent to the statement that quantum electrodynamics has a U(1) gauge symmetry.

In the strand model of quantum mechanics, potentials are introduced in precisely the same way as in usual quantum mechanics, so that the full Schrödinger equation for charged particles in external fields is recovered:

$$(i\hbar\partial_t - qV)\psi = \frac{1}{2m}(-i\hbar\boldsymbol{\nabla} - q\boldsymbol{A})^2\psi \,. \tag{141}$$

This equation is the simplest formulation of quantum theory; it describes and explains the size of atoms and molecules and thus of all objects around us, and also the colours of all things. The equation also explains interference, tunnelling, and Heisenberg's indeterminacy relations.

In summary, a non-relativistic fluctuating tangle reproduces the Schrödinger equation. An obvious question is: how does the strand model explain the influence of interactions on the rotation speed and on the wavelength of tangles? In other words: why do strands imply minimal coupling? We will answer this question in the chapter on gauge interactions. Page 207

QUANTUM INTERFERENCE FROM TANGLES

The observation of *interference* of quantum particles is due to the linear combination of states with different phases at the same position in space. Tangle functions, being wave functions, reproduce the effect. But again, it is both more fun and more instructive to explain and visualize interference with tangles.

Page 179 As mentioned above, a pure change of phase of a state ψ is defined by multiplication by a complex number of unit norm, such as $e^{i\beta}$. This corresponds to a rotation of the tangle core by an angle 2β, where the factor 2 is due to the belt trick of Figure 20. Page 168

To deduce interference, we simply use the above definition for linear combinations of tangles. This leads to the result shown in Figure 31. We find, for example, that a symmetric sum of a tangle and the same tangle with the phase rotated by $\pi/2$ (thus a core rotated by π) results in a tangle whose phase is rotated by the intermediate angle, thus $\pi/4$.

The most interesting case of interference is that of *extinction*. Scalar multiplication of a tangle function ψ by -1 gives the negative of the tangle function, the additive inverse $-\psi$. The sum of a tangle function with its negative is zero. This gives extinction in usual quantum theory. Let us check the result in the strand model, using the tangle definition of linear combinations. We have seen above that the negative of a tangle is a tangle whose core is rotated by 2π. Using the tangle definition of linear combination, we find that it is *topologically impossible* to draw or construct a localized tangle for the sum of a quantum state with its negative. The resulting tangle therefore must have vanishing crossing density in spatial regions where this operation is attempted. In short, tangles explain extinction. And as expected from quantum particles, the explanation of extinction directly involves the tangle structure.

DEDUCING THE PAULI EQUATION FROM TANGLES

As we have seen, the Schrödinger equation describes the motion of quantum particles when their spin is neglected, by assuming that spin is constant over space and time. The next step is thus to include the variations of spin over space and time. This turns out to

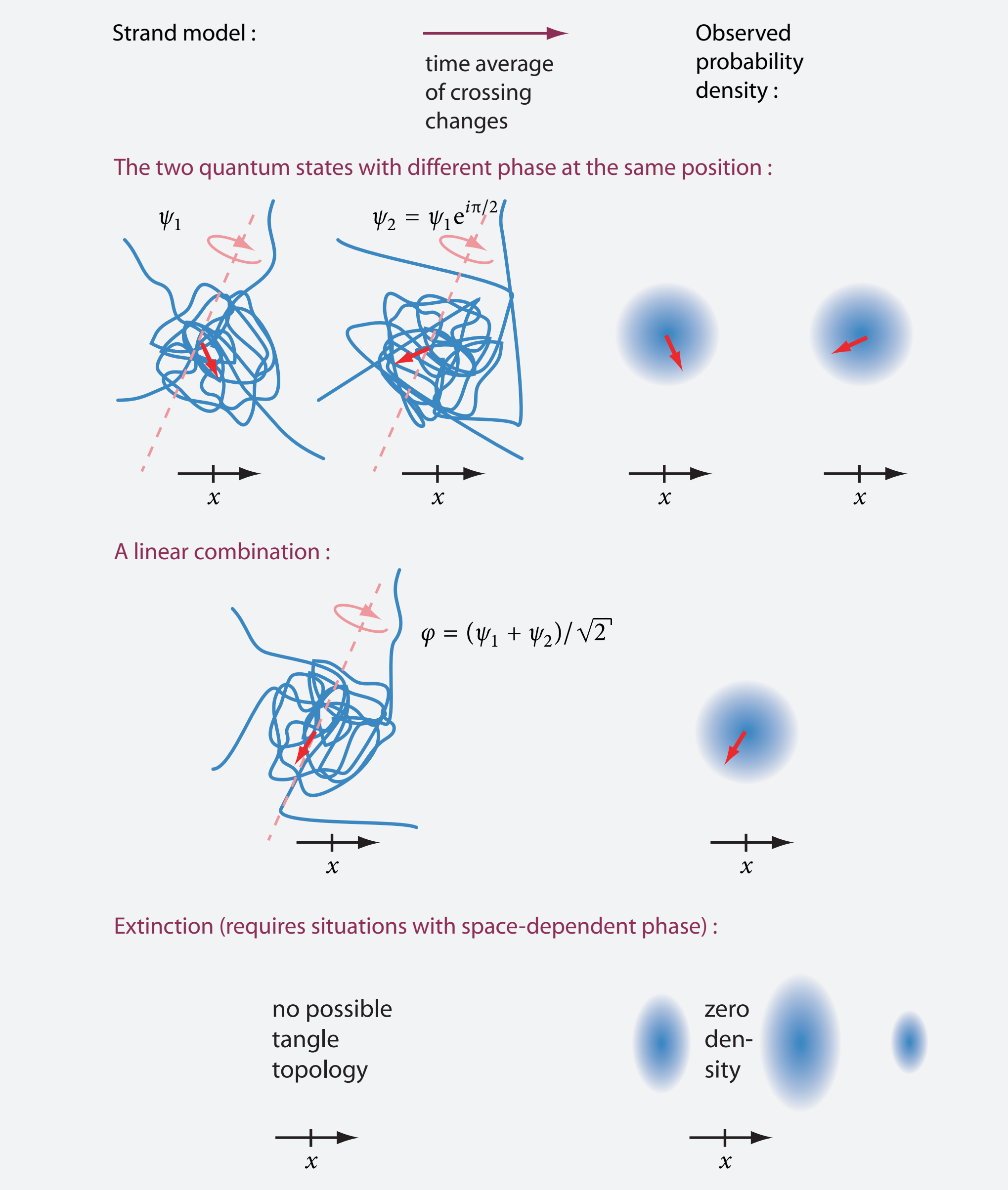


FIGURE 31 Interference: the linear combination of strands with different phase, but located at the same position.

be quite straightforward.

In the strand model, spin is modelled by the continuous rotation of a tangle. We also saw that we get wave functions from tangles if we average over short time scales. On a given position in space, a tangle function will have a local average density of crossings, a local average phase, and new, a local average orientation of the rotation axis of the tangle.

To describe the axis and orientation of the tangle core, we use the *Euler angles* α, β

Ref. 164 and γ. This yields a description of the tangle function as

$$\Psi(x,t) = \sqrt{\rho}\,\mathrm{e}^{i\alpha/2} \begin{pmatrix} \cos(\beta/2)\mathrm{e}^{i\gamma/2} \\ i\sin(\beta/2)\mathrm{e}^{-i\gamma/2} \end{pmatrix} , \qquad (142)$$

which is the natural description of a tangle that includes the orientation of the axis. As before, the crossing density is the square root of the probability density $\rho(x,t)$. The angle α, as before, describes the phase, i.e., (one half of) the rotation *around* the axis. The local orientation of the axis is described by a two-component matrix and uses the two angles β and γ. Due to the belt trick, the expression for the tangle function only contains half angles. And indeed, due to the half angles, the two-component matrix is not a vector, but a *spinor*. (The term 'spinor' was coined by the Austrian-Dutch physicist Paul Ehrenfest in analogy to 'vector' and 'tensor'. The English pronunciation is 'spinnor'.)

The other ingredient we need is a description of the spinning motion of the tangle. In contrast to the Schrödinger case, the spinning motion itself must be added in the description. A spinning tangle implies that the propagation of the wave is described by the wave vector $\boldsymbol{k}$ multiplied with the spin operator $\boldsymbol{\sigma}$. The *spin operator* $\boldsymbol{\sigma}$, for the wave function just given, is defined as the vector of three matrices

$$\boldsymbol{\sigma} = \left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right) . \qquad (143)$$

The three matrices are the well-known *Pauli matrices*.

We now take the description of the axis orientation and the description of the spinning and insert both, as we did for the Schrödinger equation, into the non-relativistic dispersion relation $\hbar\omega = E = p^2/2m = \hbar^2 k^2/2m$. We then get the wave equation

$$i\hbar\partial_t\Psi = -\frac{\hbar^2}{2m}(\boldsymbol{\sigma}\nabla)^2\Psi . \qquad (144)$$

This is *Pauli's equation* for the evolution of a free quantum particle with spin 1/2.

As final step, we include the electric and the magnetic potentials, as we did in the case of the Schrödinger equation. We again use *minimal coupling*, substituting $i\hbar\partial_t$ by $i\hbar\partial_t - qV$ and $-i\hbar\nabla$ by $-i\hbar\nabla - q\boldsymbol{A}$, thus introducing electric charge q and the potentials

Ref. 165 V and $\boldsymbol{A}$. A bit of algebra involving the spin operator then leads to the famous complete form of the Pauli equation

$$(i\hbar\partial_t - qV)\Psi = \frac{1}{2m}(-i\hbar\nabla - q\boldsymbol{A})^2\Psi - \frac{q\hbar}{2m}\boldsymbol{\sigma}\boldsymbol{B}\Psi , \qquad (145)$$

where now the magnetic field $\boldsymbol{B} = \nabla \times \boldsymbol{A}$ appears explicitly. The equation is famous for describing, among others, the motion of silver atoms, which have spin 1/2, in the Stern–

Vol. IV, page 70 Gerlach experiment. This is due to the new, last term on the right hand side, which does not appear in the Schrödinger equation. The new term is a pure spin effect and predicts a g-factor of 2. Depending on the spin orientation, the sign of the last term is either positive or negative; the term thus acts as a spin-dependent potential. The two options for the spin orientation then produce the upper and the lower beams of silver atoms that are observed in the Stern–Gerlach experiment.

In summary, a non-relativistic tangle that rotates continuously reproduces the Pauli equation. In particular, such a tangle predicts that the g-factor of an elementary charged fermion is 2.

Another simple way to visualize the equivalence between the tangle model and the Pauli equation uses the formulation of quantum theory with path integrals. We recall that the tails are not observable, and that the tangle core defines the position and phase of the quantum particle. Let us assume that the motion of the core describes the path of the particle. In this case, the continuous rotation of the central core corresponds to Ref. 166 Feynman's rotating little arrow in his famous popular book on QED. Because of its tails, the tangle obeys spinor statistics and spinor rotation behaviour. This leads to the correct interference behaviour for spin 1/2 particles. Quantum theory is then the result of the interference of all possible paths of the tangle core. In short, the path integral description of quantum theory follows directly from the description of particles as rotating tangles.

MEASUREMENTS AND WAVE FUNCTION COLLAPSE

In nature, a measurement of a quantum system in a superposition is observed to yield one of the possible eigenvalues and to prepare the system in the corresponding eigenstate. In nature, the probability of each measurement outcome depends on the coefficient of that eigenstate in the superposition.

To put the issue into context, here is a small reminder from quantum mechanics. Every measurement apparatus shows measurement results. Thus, every measurement apparatus is a device with memory. (In short, it is classical.) All devices with memory contain baths. Thus, every measurement apparatus couples baths to the system it measures. The coupling depends on and defines the observable to be measured by the apparatus. Coupling baths to *quantum* systems leads to decoherence. Decoherence leads to probabilities and wave function collapse. In short, collapse and measurement probabilities are necessary and automatic in quantum theory.

The strand model describes the measurement process in precisely the same way as standard quantum theory; in addition, it *visualizes* the process.

▷ A ***measurement*** is modelled as a strand deformation induced by the measurement apparatus that 'pulls' a tangle towards the resulting eigenstate.

▷ This pulling of strands models and visualizes the ***collapse*** of the wave function.

An example of measurement is illustrated in Figure 32. When a measurement is performed on a superposition, *the untangled 'addition region' can be imagined to shrink into disappearance.* For this to happen, one of the underlying eigenstates has to 'eat up' the other: that is the collapse of the wave function. In the example of the figure, the addition region can disappear either towards the outside or towards the inside. The choice is due to the bath that is coupled to the system during measurement; the bath thus determines the outcome of the measurement. We also conclude that the probability of measuring a particular eigenstate will depend on the (weighed) volume that the eigenstate took up in

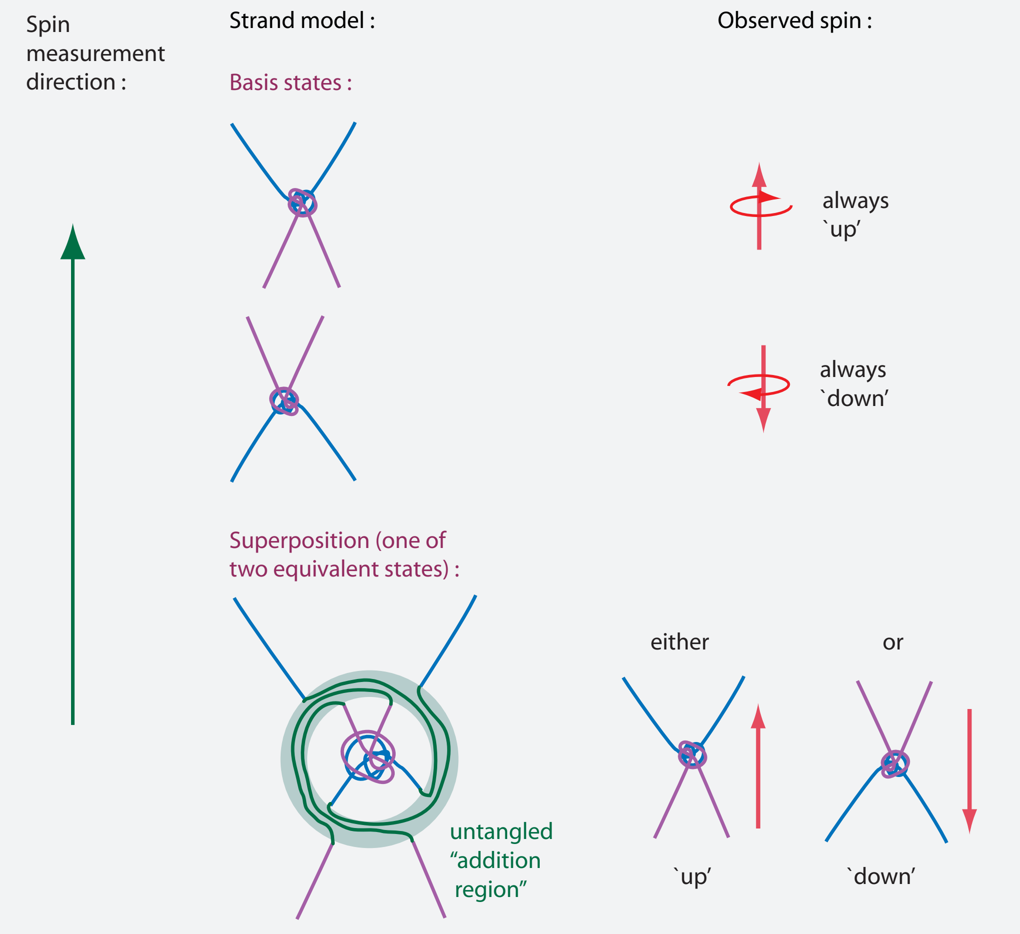


FIGURE 32 Measurement of a spin superposition: the addition region disappears either outwards or inwards.

the superposition.

This visualization of the wave function collapse also makes clear that the collapse is not limited by any speed limit, as no energy and no information is transported. Indeed, the collapse happens by displacing crossings, but does not produce any crossing changes.

In summary, the strand model describes measurements in precisely the same way as usual quantum theory. In addition, *strands visualize the collapse of the wave function as a shape deformation from a superposed tangle to a basis tangle.*

MANY-PARTICLE STATES AND ENTANGLEMENT

In nature, the quantum states of two or more particles can be *entangled*. Entangled states are multiparticle states that are not separable. Entangled states are one of the most fascinating quantum phenomena; especially in the case of macroscopic entanglement, they are still being explored in many experiments. We will discover that the strand model visualizes them simply and clearly.

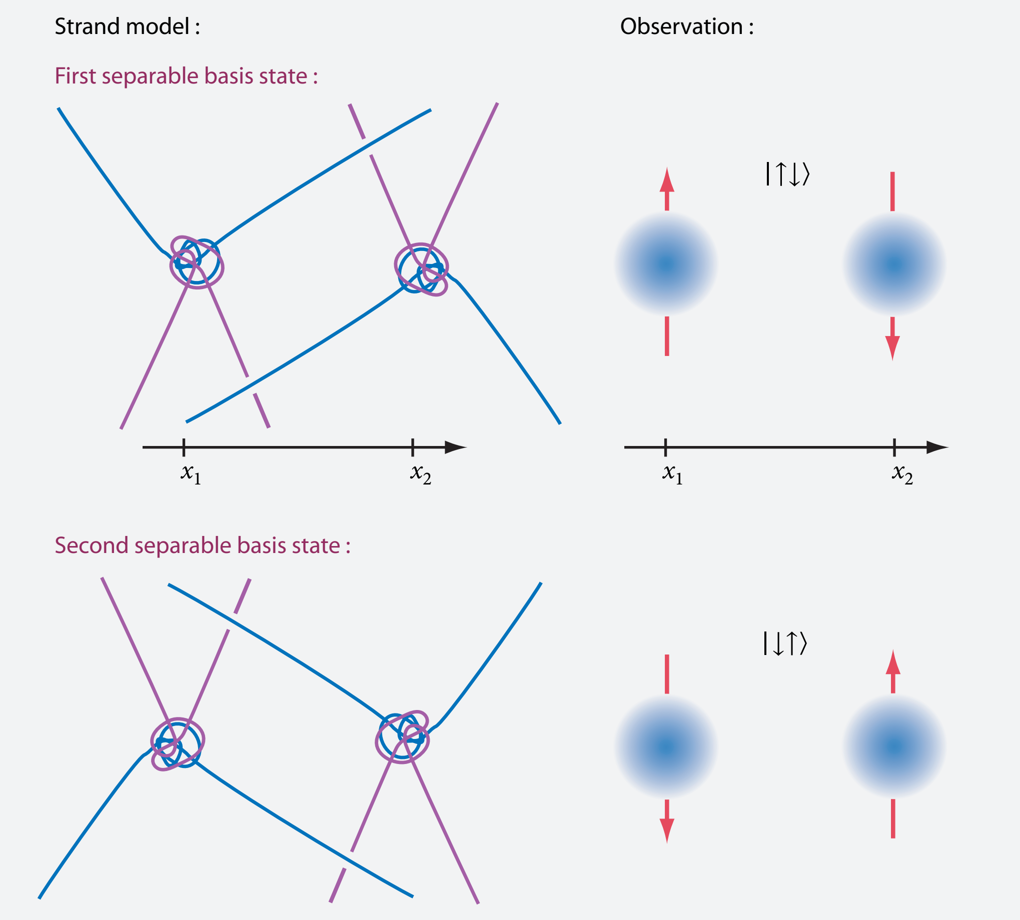


FIGURE 33 Two examples of two distant particles with spin in separable states: observation and strand model.

To describe entanglement, we need to clarify the notion of multi-particle state. In the strand model,

▷ A ***many-particle state*** is composed of several tangles.

In this way, a N-particle wave function defines N wave function values at each point in space, one value for each particle. It is well-known that in this way, many-particle Ref. 167 quantum states can be described in 3+1 dimensions. Many incorrect statements on this topic are found in the literature; many claim incorrectly the impossibility to do this and claim incorrectly the necessity of a 3N-dimensional space for N-particle wave functions. The main point is that for N particles, the strand model allows to define N wave function values at each point in space: each particle has its own tangle, and each tangle yields, by short-term averaging, a value for its wave function. The strand model thus allows to model multi-particle wave functions in three-dimensional space.

Every entangled state is a superposition of separable multi-particle states. We will

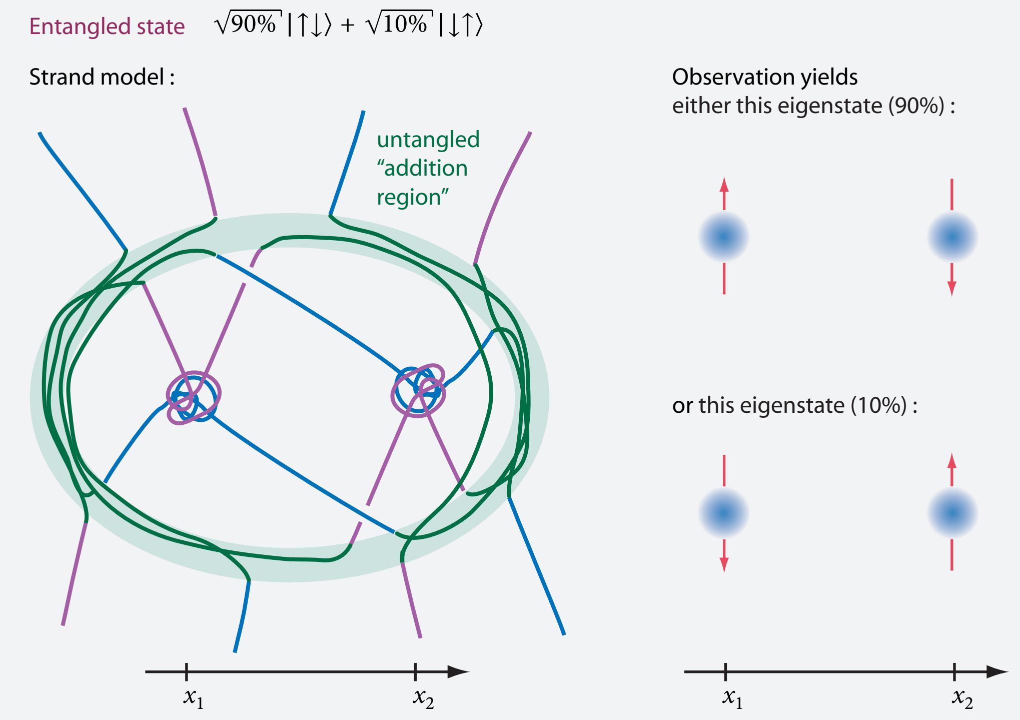


FIGURE 34 An entangled spin state of two distant particles.

now show that the definitions of superpositions and of measurement using strands are sufficient to describe entanglement.

As first example, we explore entangled states of the spin of two distant massive fermions. This is the famous example proposed by Bohm. In the strand model, two distant particles with spin 1/2 in a *separable* state are model as two distant, separate tangles of identical topology. Figure 33 shows two separable basis states, namely the two states with total spin 0 given by $|\uparrow\downarrow\rangle$ and by $|\downarrow\uparrow\rangle$. Such states can also be produced in experiments. We note that to ensure total spin 0, the tails must be imagined to cross somewhere, as shown in the figure.

We can now draw a superposition $\sqrt{90\%}\,|\uparrow\downarrow\rangle + \sqrt{10\%}\,|\downarrow\uparrow\rangle$ of the two spin-0 basis states. We simply use the definition of addition and find the state shown in Figure 34. We can now use the definition of measurement to check that the state is indeed entangled. If we measure the spin orientation of one of the particles, the untangled addition region disappears. The result of the measurement will be either the state on the inside of the addition region or the state on the outside. And since the tails of the two particles are linked, after the measurement, independently of the outcome, the spin of the two particles will always point in opposite directions. This happens for every particle distance. Despite this extremely rapid and apparently superluminal collapse, no energy travels faster than light. The strand model thus reproduces exactly the observed behaviour of entangled spin 1/2 states.

A second example is the entanglement of two photons, the well-known Aspect exper-

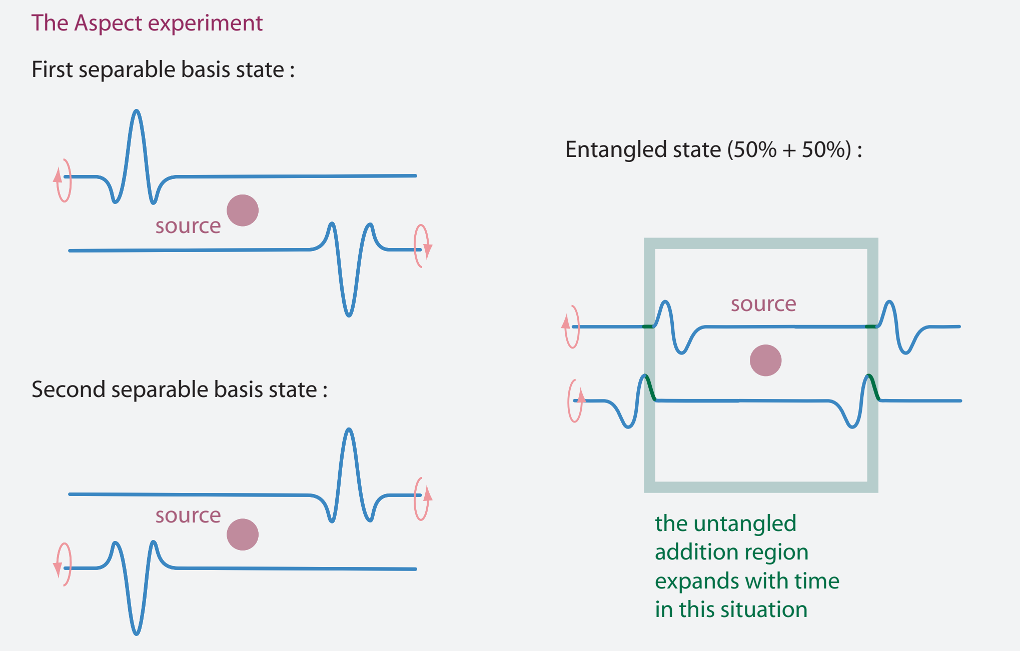


FIGURE 35 The basis states and an entangled state of two distant photons travelling in opposite directions, with total spin 0.

iment. Also in this case, entangled spin 0 states, i.e., entangled states of photons of opposite helicity (spin), are most interesting. Again, the strand model helps to visualize the situation. We use the strand model for the photon that we will deduce later on. Figure 35 shows the strand model of the two separable basis states and the strand model of the entangled state. Again, the measurement of the helicity of one photon in the entangled state will lead to one of the two basis states. And as soon as the helicity of one photon is measured, the helicity of its companion collapses to the opposite value, whatever the distance! Experimentally, the effect has been observed for distances of many kilometres. Again, despite the extremely rapid collapse, no energy travels faster than light. And again, the strand model completely reproduces the observations.

Page 212

Ref. 168

Mixed states

Mixed states are statistical ensembles of pure states. In the strand model,

▷ A ***mixed state*** is a (weighted) temporal alternation of pure states.

Mixed states are important in discussions of thermodynamic quantities. We mention them to complete the equivalence of the states that appear in quantum theory with those provided by the strand model. We do not pursue this topic any further.

The dimensionality of space-time

'Nature consists of particles moving in empty space.' Democritus stated this 2500 years ago. Today, we know that is a simplified description of one half of physics: it is a simplified description of quantum theory. In fact, Democritus' statement, together with strands, allows us to show that physical space has three dimensions, as we argue now.

Deducing the dimensionality of physical space from first principles is an old and difficult problem. Our exploration of the foundations of the strand model has shown that humans, animals and machines always use three spatial dimensions to describe their environment. They cannot do otherwise. Humans, animals and machines cannot talk and think without three dimensions as background space.

Page 162

But how can we show that *physical space* – not the *background space* we need for thinking – is really three-dimensional? We need to show that (1) all experiments reproduce the result and that (2) no other number of dimensions yields a consistent description of nature.

In nature, and also in the strand model, as long as particles can be defined, they can be rotated around each other and they can be exchanged. No experiment has ever been performed or has ever been proposed that changes this observation. The observed properties of rotations, of spin 1/2, and all other observations show that space has three dimensions. In the strand model, the position and the orientation of a particle is intrinsically a three-dimensional quantity, and physical space is thus three-dimensional, at all energies. (The only exceptions are horizons and the Planck scale.) In short, both nature and the strand model are found to be three-dimensional at all experimentally accessible energy scales. Conversely, detecting an additional spatial dimension would directly invalidate the strand model.

In the strand model, knots and tangles are impossible to construct in physical spaces with dimensions other than three. Mathematicians can show that in four spatial dimensions, every knot and every tangle can be undone. (Time is not and does not count as a fourth spatial dimension, and strands are assumed to remain one-dimensional entities in this argument.) Worse, in the strand model, spin does not exist in spaces that have more or fewer than three dimensions. In short, the strand model of matter and of observers, be they animals, people or machines, is possible only in three spatial dimensions. No description of nature with a physical space of more or less than three dimensions is possible with strands. Conversely, constructing such a description would invalidate the strand model.

To complete the consistency check, we need to explain how physical space, or the quantum vacuum, is produced by strands. In quantum field theory, the vacuum is defined as the state that contains no real, but only virtual particles. Now, the strand model shows that fluctuating strands automatically yield virtual particle–antiparticle pairs. These pairs are short-lived, but share with real particles and real antiparticles an essential property: their spin value. In other words, the strand model describes the vacuum as that state which contains only short-lived particle tangles with spin. This implies that the quantum vacuum, and thus the physical space produced by strands, has three spatial dimensions. Similar arguments also imply that physical space is homogeneous and isotropic.

Page 219

The same type of arguments can be collected for the one-dimensionality of physical

Challenge 123 e time. It can be fun exploring them – for a short while. In summary, the strand model *only* works in $3 + 1$ space-time dimensions; it does not allow any other number of dimensions. Page 158 We have thus ticked off another of the millennium issues. We can thus continue with our adventure.

Operators and the Heisenberg picture

In quantum theory, Hermitean operators play an important role. In the strand model, *hermitean* or self-adjoint operators are operators that leave the tangle topology invariant. Also unitary operators play an important role in quantum theory. In the strand model, *unitary* operators are operators that deform tangles by preserving their size and shape.

Physicists know two ways to describe quantum theory. One is to describe evolution with time-dependent quantum states – the *Schrödinger picture* we are using here – and the other is to describe evolution with time-dependent operators. In this so-called *Heisenberg picture*, the time evolution is described by the operators.

The two pictures of quantum theory are equivalent. In the Heisenberg picture, the fundamental principle, the equivalence of a crossing switch with $\hbar$, becomes a statement Ref. 169 on the behaviour of operators. Already in 1987, Louis Kauffman had argued that the commutation relation for the momentum and position operators

$$px - xp = \hbar i \tag{146}$$

is related to a crossing switch. The present chapter is the confirmation of that speculation.

In quantum mechanics, the commutation relation follows from the definition of the momentum operator as $p = \hbar k$, $k = -i\partial_x$ being the wave vector operator. The factor $\hbar$ defines the unit of momentum. The wave vector counts the number of wave crests of a wave. Now, in the strand model, a rotation of a state by an angle π is described by a multiplication by i. Counting wave crests is only possible by using the factor i, as this factor is the only property that distinguishes a crest from a trough. In short, the commutation relation follows from the fundamental principle of the strand model.

Hidden variables and the Kochen–Specker theorem

At first sight, the strand model seems to fall into the trap of introducing hidden variables into quantum theory. One could indeed argue that the shapes (and fluctuations) of the strands play the role of hidden variables. On the other hand, it is well known that non-contextual hidden variables are impossible in quantum theory, as shown by the Kochen–Ref. 170 Specker theorem (for sufficiently high Hilbert-space dimensions). Is the strand model flawed? No.

We recall that strands are not observable. In particular, strand shapes are not physical observables. Thus strand shapes are not variables. But even if we promoted strand shapes to physical variables, we notice that the evolution of the strand shapes is observable only through the ensuing crossing switches. And crossing switches evolve due to the influence of the environment, which consists of all other strands in nature, including those of space-time itself. The evolution of crossing switches is thus contextual – and so is that of the strand shapes. Therefore, the strand model does not contradict the Kochen–Specker theorem.

In short, the strand model contains no (observable) variables apart from the usual ones from quantum theory, contains no hidden variables, and in particular contains no non-contextual hidden variables. Strands lead to a contextual and probabilistic description of nature. In summary, despite using fluctuating tangles as underlying structure, the strand model is equivalent to usual quantum theory. The strand model contains nothing more and nothing less than usual quantum theory.

LAGRANGIANS AND THE PRINCIPLE OF LEAST ACTION

Before we derive the Dirac equation, we show that the strand model naturally leads to describe motion with Lagrangians.

In nature, physical action is an observable measured in multiples of the natural unit, the quantum of action $\hbar$. Action is the fundamental observable about nature, because *action measures the total change occurring in a process.*

In the strand model,

▷ The physical ***action*** W of a physical process is the observed number of crossing switches of strands. Action is a multiple of $\hbar$.

We note that these multiples do not need to be integer multiples. We further note that through this definition, *action is observer-invariant.* This important property is thus automatic in the strand model.

In nature, energy is action per time. Thus, in the strand model we have:

▷ ***Energy*** is the number of crossing switches per time in a system.

In nature, when free quantum particles move, their phase changes linearly with time. In other words, the 'little arrow' representing the free particle phase rotates with constant angular frequency. We saw that in the strand model, the 'little arrow' is taken as (half) the orientation angle of the tangle core, and the arrow rotation is (half) the rotation of the tangle core.

▷ The ***kinetic energy*** T of a particle is the number of crossing switches per time induced by shape fluctuations of the continuously rotating tangle core.

We call $\mathcal{T}$ the corresponding volume density: $\mathcal{T} = T/V$. In nature, the Lagrangian is a practical quantity to describe motion. For a *free* particle, the Lagrangian density $\mathcal{L} = \mathcal{T}$ is simply the kinetic energy density, and the action $W = \int \mathcal{L}\mathrm{d}t = \mathcal{T}t$ is the product of kinetic energy and time. In the strand model, a free particle is a constantly rotating and advancing tangle. We see directly that this constant evolution minimizes the action W for a particle, given the states at the start and at the end.

This aspect is more interesting for particles that interact. Interactions can be described by a potential energy U, which is, more properly speaking, the energy of the field that produces the interaction. In the strand model,

▷ ***Potential energy*** U is the number of crossing switches per time induced by an interaction field.

We call $\mathcal{U}$ the corresponding volume density: $\mathcal{U} = U/V$. In short, in the strand model, an interaction changes the rotation rate and the linear motion of a particle tangle.

In the strand model, the *difference* between kinetic and potential energy is thus a quantity that describes how much a system consisting of a tangle and a field *changes* at a given time. The total change is the integral over time of all instantaneous changes. In other words, in the strand model we have:

- The ***Lagrangian density*** $\mathcal{L} = \mathcal{T} - \mathcal{U}$ is the number of crossing switches per volume and time, averaged over many Planck scales.
- The physical ***action*** $W = \int L\,\mathrm{d}t = \int \mathcal{L}\,\mathrm{d}V\mathrm{d}t$ of a physical process is the observed number of crossing switches of strands. The action value W_{if} between an initial state ψ_{i} and a final state ψ_{f} is given by

$$W_{\mathrm{if}} = \langle \psi_{\mathrm{i}} \,|\int \mathcal{L}\,\mathrm{d}t\,|\, \psi_{\mathrm{f}} \rangle = \langle \psi_{\mathrm{i}} \,|\int \mathcal{T} - \mathcal{U}\,\mathrm{d}t\,|\, \psi_{\mathrm{f}} \rangle \;. \tag{147}$$

Since energy is related to crossing switches, it is natural that strand fluctuations that do *not* induce crossing switches are *favoured*. In short, the strand model states

- Evolution of tangles **minimizes the action** W.

In the strand model, *the least action principle appears naturally*. In the strand model, an evolution has least action when it occurs with the smallest number of crossing changes. One can also show that the strand model implies Schwinger's quantum action principle.

Challenge 124 ny

To calculate quantum motion with the principle of least action, we need to define the kinetic and the potential energy in terms of strands. There are various possibilities for Lagrangian densities for a given evolution equation; however, all are equivalent. In case of the free Schrödinger equation, one possibility is:

$$\mathcal{L} = \frac{\mathrm{i}\hbar}{2}(\overline{\psi}\,\partial_t\psi - \partial_t\overline{\psi}\,\psi) - \frac{\hbar^2}{2m}\boldsymbol{\nabla}\overline{\psi}\boldsymbol{\nabla}\psi \tag{148}$$

In this way, the principle of least action can be used to describe the evolution of the Schrödinger equation. The same is possible for situations with potentials, for the Pauli equation, and for all other evolution equations of quantum particles.

SPECIAL RELATIVITY: THE VACUUM

In nature, there is an invariant limit energy speed c, namely the speed of light and of all other massless radiation. Special relativity is the description of the consequences from this observation, in the case of a flat space-time.

We remark that special relativity also implies and requires that the flat vacuum looks the same for all inertial observers. In the strand model, the idea of flat vacuum as a set of

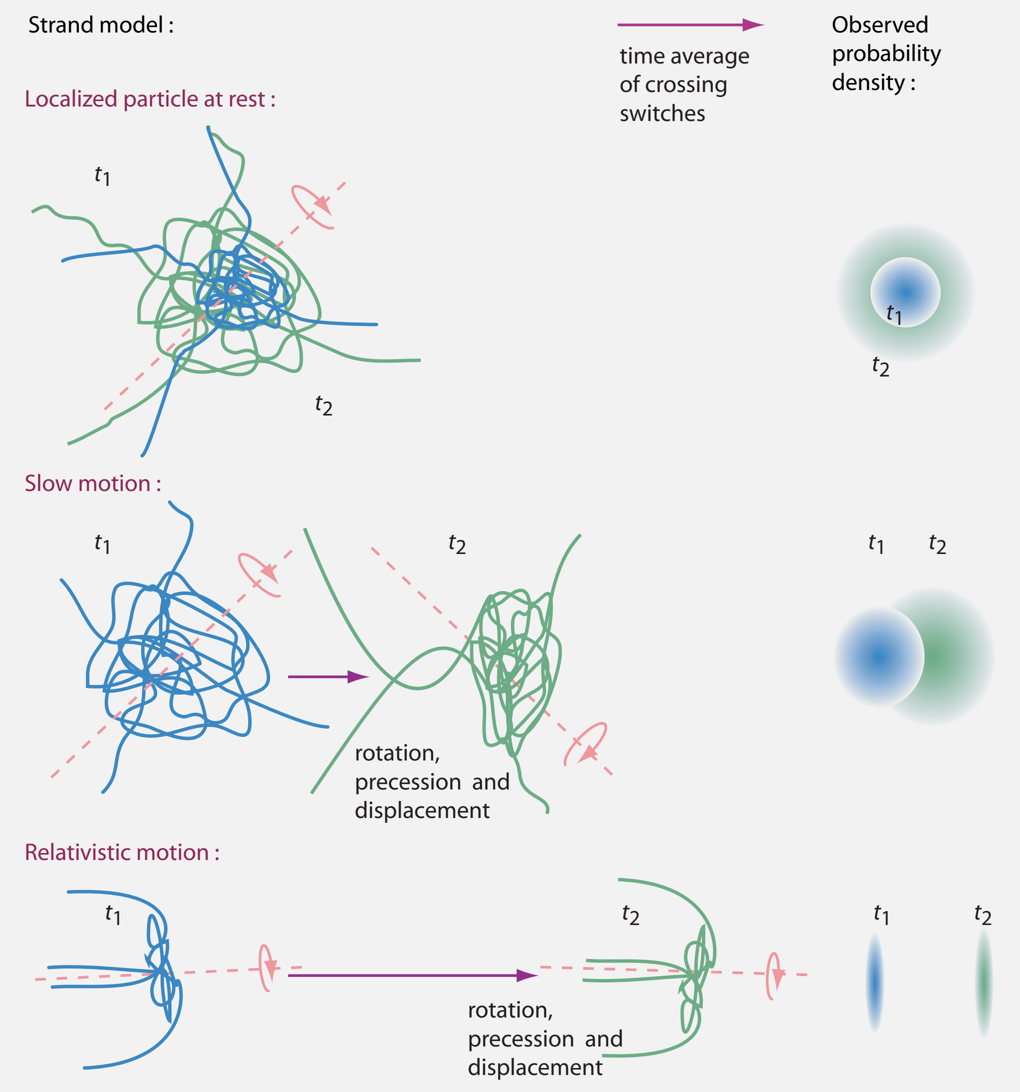


FIGURE 36 Tangles at rest, at low speed and at relativistic speed.

fluctuating featureless strands that are *unknotted* and *unlinked* automatically implies that for any inertial observer the flat vacuum has no matter content, has no energy content, is isotropic and is homogeneous. The strand model thus realizes this basic requirement of special relativity.

We note that in the strand model, the vacuum is unique, and the vacuum energy of flat infinite vacuum is exactly zero. In the strand model, there is no divergence of the vacuum energy, and there is thus *no* contribution to the cosmological constant from quantum field theory.

SPECIAL RELATIVITY: THE INVARIANT LIMIT SPEED

In the strand model, massless particles are unknotted and untangled. Even though we will deduce the strand model for photons only later on, we use it here already, to speed up the discussion. In the strand model, the *photon* is described a single, helically deformed unknotted strand, as shown in Figure 45. Therefore, we can define: Page 212

- ▷ The ***Planck speed*** c is the observed average speed of crossing switches due to photons.

Because the definition uses crossing switches, the limit speed c is an *energy speed*. The speed of light c is an average. Indeed, as is well-known in quantum field theory, single photons can travel faster or slower than light, but the probability for large deviations is extremely low. Ref. 171

The linear motion of a helically deformed strand through the strands that make up the vacuum can thus be imagined like the motion of a bottle opener through cork. The strands of the vacuum do not need to fluctuate to allow the motion of a photon. In contrast, the linear motion of a matter tangle through vacuum involves the vacuum strands and their fluctuations. This results in an effective slowing down of the motion of matter tangles through vacuum. We have thus deduced that matter tangles always move *more slowly than light*.

In fact, we see that ultrarelativistic tangles move, as shown in Figure 36, almost like light. We thus find that matter can *almost* reach the speed of light. The speed c is thus a *limit speed*.

The speed c is defined as an average, because, as well-known in quantum field theory, there are small probabilities that light moves faster or slower that c. But the average result c will be the same for every observer. The value of the speed c is thus *invariant*.

In 1905, Einstein showed that the three mentioned properties of the speed of light – energy speed, limit speed, and invariant speed – imply the Lorentz transformations. In particular, the three properties of the speed of light c imply that the energy E of a particle of mass m is related to its momentum p as

$$E^2 = m^2 c^4 + c^2 p^2 \quad . \qquad (149)$$

This dispersion relation is thus also valid for massive particles made of tangled strands – even though we cannot yet calculate tangle masses. (We will do this later on.) Page 306

Page 150 Should we be surprised at this result? No. In the fundamental principle, the definition of the crossing switch, we inserted the speed of light as the ratio between the Planck length and the Planck time. Therefore, by defining the crossing switch in the way we did, we have implicitly stated the invariance of the speed of light.

Fluctuating strands imply that flat vacuum has no matter or energy content, for *every* inertial observer. Due to the strand fluctuations, flat vacuum is also homogeneous and isotropic for every inertial observer. Therefore, together with the 3 + 1-dimensionality of space-time deduced above, we have now shown that flat vacuum has Poincaré symmetry. Page 192
This settles another issue from the millennium list. Page 158

However, one problem remains open: how exactly do tangles move through the web that describes the vacuum? We will leave this issue open for a while, and come back to it later on.

Page 302

Page 182

The relativistic dispersion relation differs from the non-relativistic case in two ways. First, the energy scale is shifted, and now includes the rest energy $E_0 = mc^2$; secondly, the spin precession is not independent of the particle speed any more. For ultrarelativistic particles, the spin lies in the direction of the motion.

If we neglect spin, we can use the relativistic dispersion relation to deduce directly the well-known Schrödinger–Klein–Gordon equation for the evolution of a wave function:

$$-\hbar^2 \partial_{tt}\psi = m^2c^4 - c^2\hbar^2\nabla^2\psi \;. \tag{150}$$

In other words, the strand model implies that relativistic tangles follow the Schrödinger–Klein–Gordon equation. We now build on this result to deduce Dirac's equation for relativistic quantum motion.

DIRAC'S EQUATION DEDUCED FROM TANGLES

The relativistic Schrödinger–Klein–Gordon equation assumes that spin is negligible. This approximation fails to describe most experiments. A precise description of relativistic elementary particles must include spin.

We deduced the Schrödinger equation using the relation between phase and the quantum of action, the non-relativistic energy–momentum relation, and neglecting spin. In the next step we deduced the Pauli equation using again the relation between phase and the quantum of action, again the non-relativistic energy–momentum relation, but now using the properties of spin 1/2. The following step was to deduce the Schrödinger–Klein–Gordon equation using again the relation between phase and the quantum of action, this time the relativistic energy–momentum relation, but assuming zero spin. The final and correct description of elementary fermions, the Dirac equation, results from combining all three ingredients: (1) the relation between the quantum of action and the phase of the wave function, (2) the relativistic mass–energy relation, and (3) spin 1/2. We can do this because all three ingredients are reproduced by the strand model.

The main observation about spin in the relativistic context is that there are states of right-handed and of left-handed chirality: the spin can precess in two opposite senses around the direction of momentum. Above all, for massive particles, the two chiral states mix. The existence of two chiralities requires a description of spinning particles with a wave function that has *four* complex components, thus *twice* the number of components that appear in the Pauli equation. Indeed, the Pauli equation assumes only one, given sign for the chirality, even though it does not specify it. This is possible because in non-relativistic situations, states of different chirality do not mix.

Consistency requires that each of the four components of the wave function of a relativistic spinning particle must follow the relativistic Schrödinger–Klein–Gordon equation. This requirement is known to be sufficient to deduce the Dirac equation. One of the simplest derivations is due to Lerner; we summarize it here.

Ref. 172

Ref. 173

When a spinning object moves relativistically, we must take both chiralities into account. We call u the negative chiral state and v the positive chiral state. Each state is

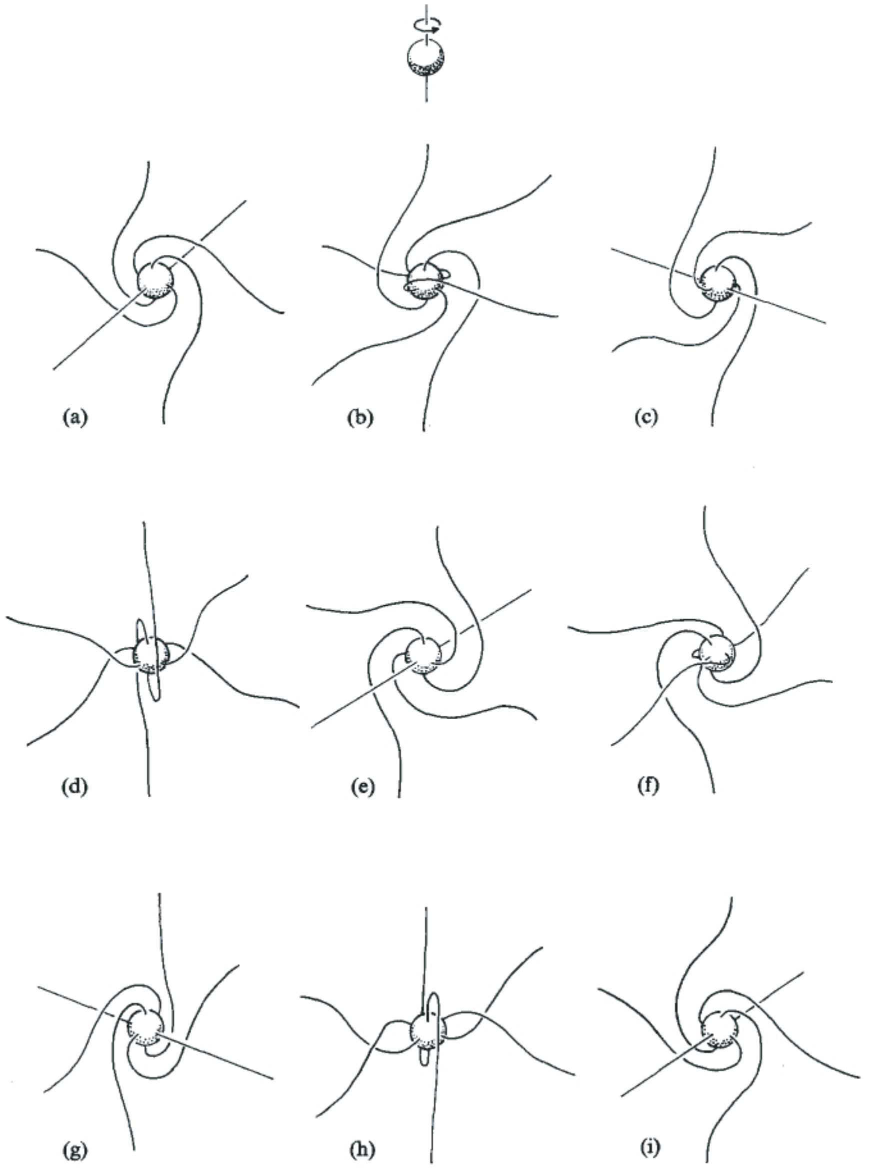


FIGURE 37 The belt trick for a rotating body with many tails, as used by Battey-Pratt and Racey to deduce the Dirac equation (© Springer Verlag, from Ref. 174)

described by two complex numbers that depend on space and time. The 4-vector for probability and current becomes

$$J_\mu = u^\dagger \sigma_\mu u + v^\dagger \sigma_\mu v \ . \tag{151}$$

We now introduce the four-component spinor φ and the 4×4 spin matrices a_μ

$$\varphi = \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{and} \quad \alpha_\mu = \begin{pmatrix} \sigma_\mu & 0 \\ 0 & \overline{\sigma}_\mu \end{pmatrix} , \tag{152}$$

where $\sigma_\mu = (I, \boldsymbol{\sigma})$ and $\overline{\sigma}_\mu = (I, -\boldsymbol{\sigma})$ and I is the 2×2 identity matrix. The 4-current can then be written as

$$J_\mu = \varphi^\dagger \alpha_\mu \varphi . \tag{153}$$

The three requirements of current conservation, Lorentz invariance and linearity then Ref. 173 yield the evolution equation

$$\mathrm{i}\hbar\partial^\mu(\alpha_\mu\varphi) + mc\gamma_5\varphi = 0 . \tag{154}$$

This is the Dirac equation in the (less usual) spinorial representation.* The last term shows that mass couples right and left chiralities. The equation can be expanded to include potentials using minimal coupling, in the same way as done above for the Schrödinger and Pauli equations.

The derivation of the Dirac equation can be repeated and visualized also with the help Ref. 174 of strands. This was done for the first time by Battey-Pratt and Racey, in 1980. They explored a central object connected by strands (or 'tails') to the border of space, as shown in Figure 37. In their approach, the central object plus the tails correspond to a microscopic particle. The central object is assumed to be continuously rotating. (In the strand model, the central object becomes the tangle core.) Battey-Pratt and Racey then explored relativistically moving objects of both chiralities. Studying the evolution of the phases and axes for the chiral objects yields the Dirac equation. The derivation by Battey-Pratt and Racey is mathematically equivalent to the one just given.

The belt trick is fundamental for understanding the Dirac equation. In the strand model, a strand can rotate in two directions, and the belt trick can occur in two directions. The resulting four combinations form the four components of the Dirac spinor and of the Dirac equation.

In summary, tangles completely reproduce both the rotation and the linear motion of elementary fermions.

VISUALIZING SPINORS AND DIRAC'S EQUATION USING TANGLES

Despite its apparent complexity, the Dirac equation makes only three statements: spin 1/2 particles are fermions, obey the relativistic energy–momentum relation, and the wave function behaves like a wave. All three statements are visualized by the tangle model of fermions: tangles behave as spinors, the relativistic energy–momentum relation is built-in, and rotating tangle cores reproduce the evolution of the phase. Let us look at the details.

---

* The matrix γ_5 is defined here as

$$\gamma_5 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} , \tag{155}$$

where I is the 2×2 identity matrix.

Given a particle tangle, the short-time fluctuations lead, after averaging, to the wave function. The tangle model of fermions provides a *visualization* of the *spinor* wave function. Indeed, at each point in space, the wave function has the following parameters:

— There is an average density $\sqrt{\rho(x,t)}$; physically, this is the square root of the probability density. In the strand model, this is the local crossing density.
— There is a set of three Euler angles α, β and γ; physically, they describe the average local orientation and phase of the spin axis. In the strand model, this is the average local orientation and phase of the tangle core.
— There is a second set of three parameters $\boldsymbol{v} = (v_x, v_y, v_z)$; physically, they describe, at one's preference, either the average local Lorentz boost or a second set of three Euler angles. In the strand model, these parameters describe the average local deformation of the core that is due to the Lorentz boost.
— There is a phase δ; physically, this represents the relative importance of particle and antiparticle density. In the strand model, this phase describes with what probability the average local belt trick is performed right-handedly or left-handedly.

In total, these are eight real parameters; they correspond to one positive real number and seven phases. They lead to the description of a spinor wave function as Ref. 175

$$\varphi = \sqrt{\rho}\ \mathrm{e}^{i\delta}\, L(\boldsymbol{v})\, R(\alpha/2, \beta/2, \gamma/2)\ , \qquad (156)$$

where the product LR is an abbreviation for the boosted and rotated unit spinor. It is equivalent to the description with four complex parameters used in most textbooks. This description of a spinor wave function and the related physical visualization of its density and its first six phases dates already from the 1960s. The visualisation can be deduced from the study of relativistic spinning tops or of relativistic fluids. Rotating tangles are more realistic, however. In contrast to all previous visualizations, the rotating tangle model explains also the last phase, i.e., the phase that describes matter and anti-matter, explains the appearance of the quantum of action $\hbar$, and explains fermion behaviour. Ref. 175

In short, only rotating tangles together with the fundamental principle provide a simple, complete and precise visualisation of spinor wave functions and their evolution. The tangle model for quantum particles is a simple extension of Feynman's idea to describe a quantum particle as a rotating little arrow. The arrow can be imagined as being attached to the rotating tangle core. The tails are needed to reproduce fermion behaviour. The specific type of tangle core determines the type of particle. The blurring of the crossings defines the wave function. Rotating arrows describe non-relativistic quantum physics; rotating tangles describe relativistic quantum physics. Ref. 166

Visualizing spinor wave functions with tangles of strands helps the understanding of the Dirac equation in several ways.

1. Tangles support the view that elementary particles are little rotating entities, also in the relativistic case. This fact has been pointed out by many scholars over the years. The strand model provides a consistent visualization for these discussions. Ref. 175
2. The belt trick can be seen as the mechanism underlying the famous Zitterbewegung that is part of the Dirac equation. The limitations in the observing the belt trick translate directly into the difficulties of observing the Zitterbewegung. Ref. 176

3. The belt trick also visualizes why the velocity operator for a relativistic particle has eigenvalues $\pm c$.
4. The Compton length is often seen as the typical length at which quantum field effects take place. In the tangle model, it would correspond to the average size needed for the belt trick. The strand model thus suggests that the mass of a particle is related to the average size needed for the belt trick.
5. Tangles support the – at first sight bizarre – picture of elementary particles as little charges rotating around a centre of mass. Indeed, in the tangle model, particle rotation requires a regular application of the belt trick of Figure 20, and the belt trick can be interpreted as inducing the rotation of a charge, defined by the tangle core, around a centre of mass, defined by the average of the core position. It can thus be helpful to use the strand model to visualize this description.

Ref. 177

Page 168

6. The tangle model can be seen as a vindication of the stochastic quantization research programme; quantum motion is the result of underlying fluctuations. For example, the similarity of the Schrödinger equation and the diffusion equation is modelled and explained by the strand model: since crossings can be rotated, diffusion of crossings leads to the imaginary unit that appears in the Schrödinger equation.

Ref. 178

In short, rotating tangles are the underlying model for the propagation of fermions. *Tangles model propagators.* This modelling is possible because the Dirac equation results from only three ingredients:

— the relation between the quantum of action and the phase of the wave function (the wave behaviour),

— the relation between the quantum of action and spinor behaviour (the exchange behaviour),

— and the mass–energy relation of special relativity (the particle behaviour).

And all three ingredients are reproduced by the strand model. The apparent complexity of the Dirac equation hides its fundamental simplicity. The tangle model reproduces the ingredients of the Dirac equation, reproduces the equation itself, and makes the simplicity manifest.

In summary, tangles can be used as a precise visualization and explanation of quantum physics. Wave functions are *blurred tangles* – with the detail that not the strands, but their crossings are blurred.

THE DIFFERENCE BETWEEN QUANTUM MECHANICS AND QUANTUM FIELD THEORY

In the strand model, *quantum mechanics* is the approximation in which a particle is described by a tangle with a topology that is *fixed* in time. This approximation allows to derive the Dirac equation, the Schrödinger–Klein–Gordon equation, the Proca equation, the Pauli equation, etc. In this approximation, the strand model for the electron in a hydrogen atom is illustrated in Figure 38. The approximation also allows to deduce the existence of the three gauge interactions, as we will see in the next chapter.

In contrast, *quantum field theory* appears when it becomes clear that particles are not tangles with fixed topology, but that for each particle, the topology *varies* inside a specific family of tangles. As we will see later on, this topology variation is necessary to describe SU(2) breaking, to explain the existence of three fermion generations, and to calculate

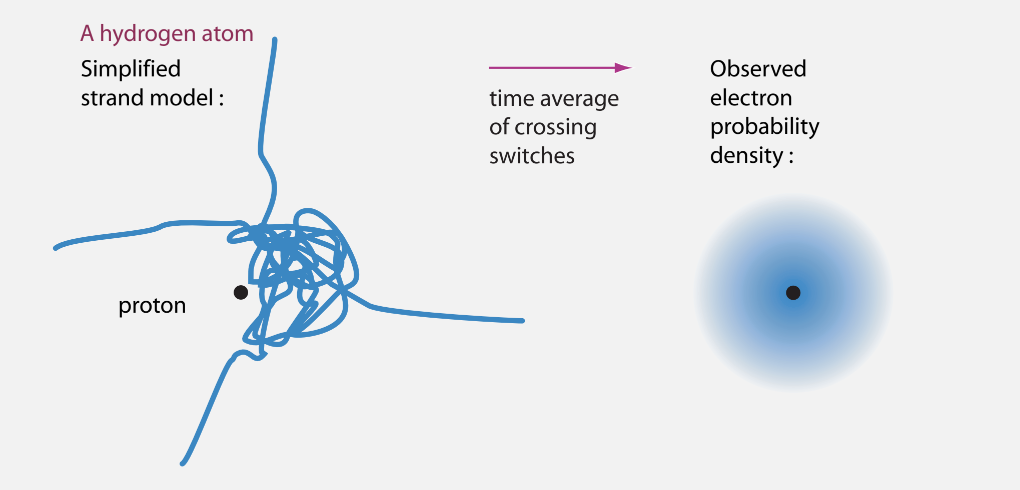


FIGURE 38 A simple, quantum-mechanical view of a hydrogen atom.

particle masses and couplings.

A FLASHBACK: SETTLING THREE PARADOXES OF GALILEAN PHYSICS

In all descriptions of physics, space and time are measured, explained and defined with matter, for example, with the help of metre bars and clocks. On the other side, matter is measured, explained and defined with space and time. This interdependence is a natural consequence of strands. Both matter and space-time turn out to be approximations of the same basic building blocks; this common origin explains the apparent circular reasoning of Galilean physics. Most of all, the strand model changes it from a paradox to a logical necessity.

The strand model defines vacuum, and thus physical space, as a result of averaging strand crossings. Space is thus a *relative* concept. Newton's bucket experiment is sometimes seen as a counterargument to this conclusion and as an argument for absolute space. However, the strand model shows that any turning object is connected to the rest of the universe through its tails. This connection makes every rotation an example of relative motion. Rotation is thus always performed relatively to the horizon of the universe. On the other hand, the detection of tangles among the tails allows a *local* determination of the rotation state, as is observed. Strands thus confirm that rotation and space are relative concepts. Strands thus also explain why we can turn ourselves on ice by rotating an arm over our head, without outside help. Strands lie to rest all issues around the rotating bucket.

A long time ago, Zeno of Elea based one of his paradoxes – the flying arrow that cannot reach the target – on an assumption that is usually taken as granted: he stated the impossibility to distinguish a short-time image (or state) of a moving body form the image (or state) of a resting body. The flattening of the tangles involved shows that the assumption is incorrect; motion and rest are *distinguishable*, even in (imagined) photographs taken with extremely short shutter times. The argument of Zeno is thus not possible, and the paradox disappears.

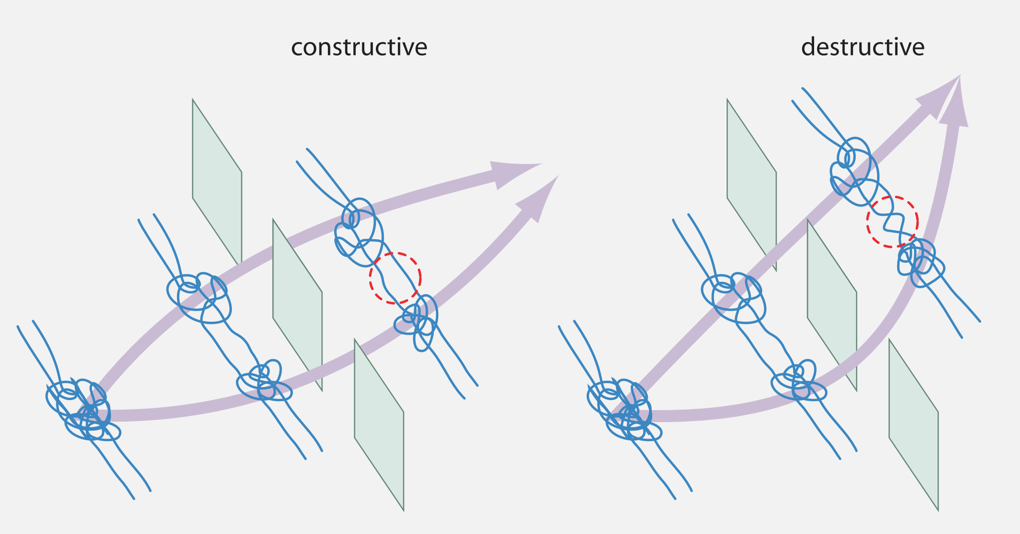


FIGURE 39 A fermion tangle passing a double slit: constructive interference (left) and destructive interference (right).

Fun challenges about quantum theory

> Urlaub ist die Fortsetzung des Familienlebens
> unter erschwerten Bedingungen.*
> Dieter Hildebrandt

Are the definitions of wave function addition and multiplication also valid for the spinor tangle functions? Challenge 125 s

* *

Interference can be visualized with strands. Fermion interference is visualized in Figure 39, photon interference in Figure 40.

* *

In the strand model, tangle energy is related to tangle rotation. What is the difference between the angular frequency for tangles in the non-relativistic and in the relativistic case? Challenge 126 ny

* *

In the strand description of quantum mechanics, strands are impenetrable: they cannot pass through each other (at finite distances). Can quantum mechanics also be derived if the model is changed and this process is allowed? Is entanglement still found? Challenge 127 s

* *

Page 171 At first sight, the apheresis machine diagram suggests that using the belt trick, animals could grow and use wheels instead of legs, because rotating wheels could be supplied

---

* 'Vacation is the continuation of family life under aggravated conditions.'

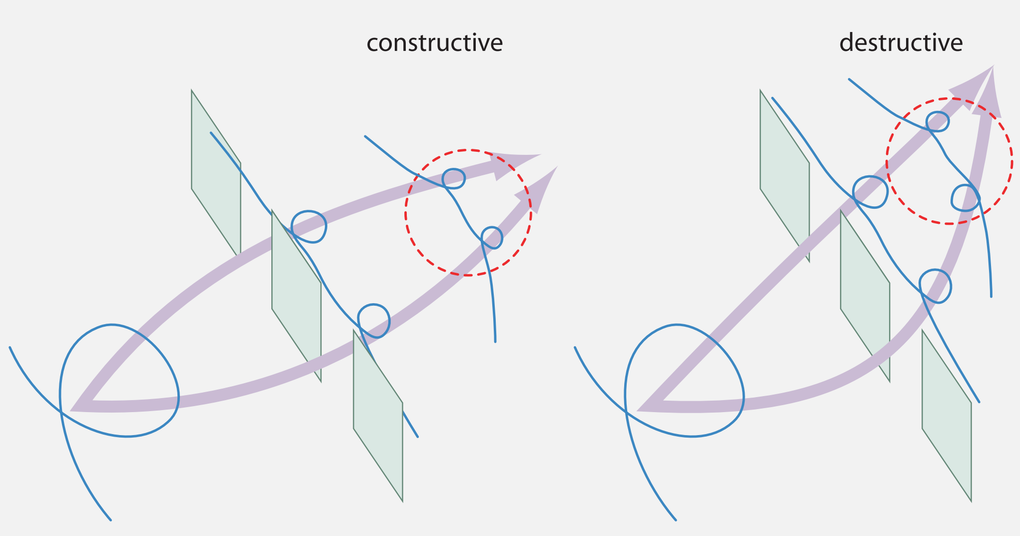


FIGURE 40 The double-slit experiment with photons: constructive interference (left) and destructive interference (right).

Challenge 128 s with blood and connected to nerves. Why did wheels not evolve nevertheless?

* *

Are the four *Wightman axioms* of quantum field theory fulfilled by the strand model without interactions? Challenge 129 ny

SUMMARY ON FERMIONS: MILLENNIUM ISSUES AND EXPERIMENTAL PREDICTIONS

In this chapter, we used the fundamental principle – crossing switches define the quantum of action $\hbar$ and the other Planck units – to deduce that particles are tangles of strands, and wave functions are time-averaged rotating tangles. In simple words, a wave function Ψ is a blurred tangle. We also deduced that blurred tangles obey the least action principle and the Dirac equation.

In other words, the strand model has confirmed Bohr's statement: quantum theory is indeed a consequence of the quantum of action. Visualizing the quantum of action as a crossing switch implies quantum theory. The strand model thus shows that all quantum effects are consequences of extension. Finally, the strand model confirms that the Dirac equation is essentially the infinitesimal version of the string trick (or belt trick).

In other words, we have shown that strands reproduce the relativistic Lagrangian density $\mathcal{L}$ of charged, elementary, relativistic fermions in an external field

$$\mathcal{L} = \overline{\varphi}(i\hbar c \not{D} - mc^2)\varphi \ , \tag{157}$$

where

$$\not{D} = \gamma^{\sigma} \mathrm{D}_{\sigma} = \gamma^{\sigma}(\partial_{\sigma} - iqA_{\sigma}) \ . \tag{158}$$

There are no known deviations between observations about elementary particles and this Lagrangian. We thus conclude that *strands reproduce the quantum theory of matter*.

We mention that the strand model predicts deviations from this Lagrangian *only* when gravity or other interactions start to play a role. For the case of gravity, the strand model predicts that deviations occur only when the energy–momentum of an elementary particle approaches the Planck value. In addition, the strand model predicts that the Planck values for momentum and energy are limit values that cannot be exceeded. So far, all experiments agree with these predictions.

Page 36

Page 158

Let us evaluate the situation. In our quest to explain the open issues of the millennium list, we have explained the origin of Planck units, the origin of wave functions, the origin of the least action principle, the origin of space-time dimensions, the Lorentz and Poincaré symmetries, the origin of particle identity, and a part of the Lagrangian of quantum field theory, namely, in particular, the Lagrangian of free fermions, such as the electron. Therefore, for the next leg, we turn to the parts of the standard model Lagrangian that are missing: those due to gauge interactions.

Chapter 10

GAUGE INTERACTIONS DEDUCED FROM STRANDS

What are interactions? At the start of this volume, when we summarized what relates the Planck units to relativity and to quantum theory, we pointed out that the nature of interactions at Planck scale was still in the dark. In the year 2000, it was known that the essential properties of the electromagnetic, the weak and the strong nuclear interaction are their respective gauge symmetries: all three interactions are *gauge interactions*. But the underlying reason for this property was unknown.

Page 17

Ref. 179

In this chapter we discover that fluctuating strands in three spatial dimensions explain the existence of precisely three interactions, each with precisely the gauge symmetry that is observed. In short, we will deduce quantum field theory from strands. In fact, strands provide a mechanism for interactions that explains and implies Feynman diagrams. The term 'mechanism' has to be taken with a grain of salt, because there is nothing mechanical involved; nevertheless, the term is not wrong, as we shall see shortly. In this chapter, we work in *flat* space-time, as is always done in quantum field theory. We leave the quantum aspects of curved space-time and of gravitation for the next chapter.

Ref. 180

Interactions and tangle core rotation

Experiments in the quantum domain show that interactions change the phase of wave functions. In the strand model, this means that *interactions rotate tangles*. But how precisely does this happen? The strand model will give us a simple answer: the emission and the absorption of gauge bosons is only possible *together* with the rotation of tangle cores. To explain this connection, we need to study the rotation of tangle cores in more detail.

When we explored spin and its connection to the belt trick, we pictured the rotation of the tangle core in the same way as the rotation of a belt buckle. The core of the tangle was rotated like a rigid object; the actual rotation occurred through the shape fluctuations of the tails. Why did we do this?

Page 169

In the strand model, *linear* particle motion is modelled as the change of position of the tangle core. It was thus obvious to model particle *spin* as the rotation of the core. We boldly drew an arrow into the cloud that described its phase and described spin as the rotation of the core with its attached arrow, as shown again in Figure 41. This bold approach led us to the Dirac equation.

However, we swept a problem under the rug: how is the phase arrow defined? How is the arrow 'attached' to the tangle core? And how does the definition, the 'attachment',

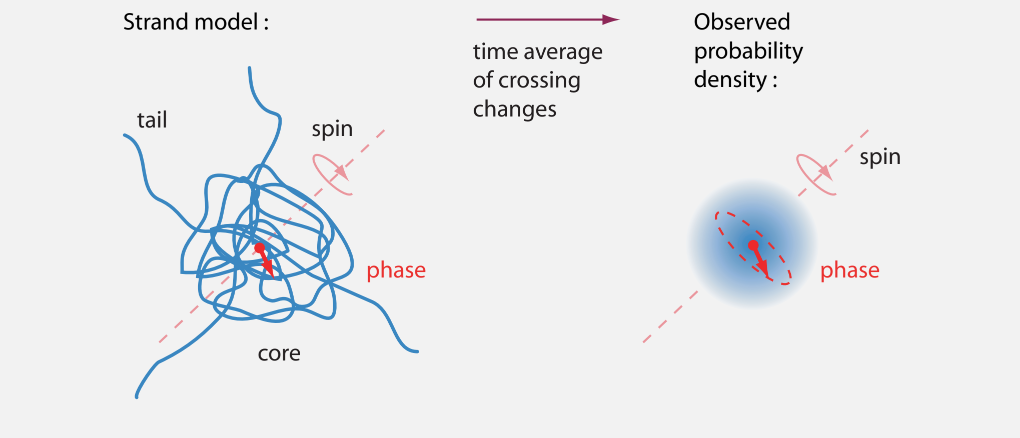


FIGURE 41 In the chapter on quantum theory, the phase was defined assuming a *rigidly rotating core*; this approximation was also used in the description of particle translation.

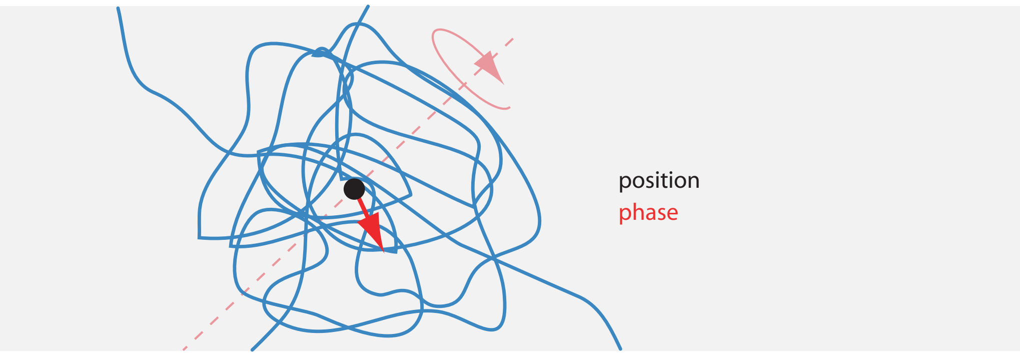


FIGURE 42 A magnified tangle core shows that the phase can also change due to *core deformations*; this leads to gauge interactions.

change with fluctuations of the core itself? It turns out that the answers to these questions automatically lead to the existence of gauge interactions; in fact, the answers automatically lead to precisely those three gauge interactions that we observe in nature.

Tail deformations versus core deformations

We can summarize the previous chapter, on the motion of matter tangles, as the chapter that focused on fluctuations of *tails*. Indeed, we have seen that

▷ ***Space-time symmetries*** are due to *tail* deformations.

Translation, rotation, boost, spin and particle exchange are all due to tail deformations; the tangle core is assumed to remain unchanged and rigid. In contrast, the present chapter focuses on fluctuations in *tangle cores*:* we will discover that

---

Ref. 181 * The contrast between tail deformations and core deformations has a *remote* similarity to gravity/gauge duality, or AdS/CFT correspondence, and to space–time duality. For example, in the strand model, the three

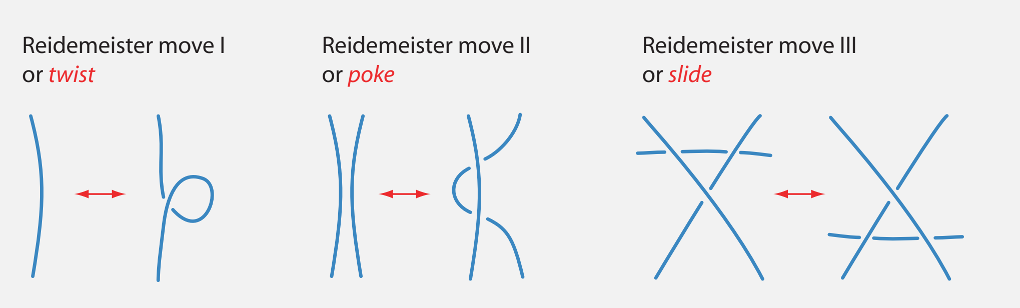


FIGURE 43 The Reidemeister moves: the three types of deformations that induce crossing switches – if the moves are properly defined.

- ▷ ***Gauge symmetries*** are due to *core* deformations.

Let us explore the tangle core in more detail. Figure 42 shows a magnified view of the core and its phase arrow. As an example, we can imagine that the phase arrow is chosen in such a way as to point to the *outermost* crossing of the rotating core. But the figure makes it clear that this definition, or any other we may use instead, implies that the phase arrow will be sensitive to the fluctuations and deformations of the strand segments that make up the core. In particular, this observation implies:

- ▷ When the phase of a core changes through *rigid orientation change*, we speak of ***rotation***.
- ▷ When the phase of a core changes through *core shape deformation*, we speak of ***interaction***.

We thus need to understand two things: First, what kinds of core deformation are there? Secondly, how precisely is the phase – i.e., each arrow definition – influenced by core deformations?

The first question, on the classification of the core deformations, is less hard than it might appear. Ref. 180 We start with our fundamental principle: events are crossing switches of strands. This implies that deformations are observable only if they induce crossing switches. Smaller or simpler deformations should not have any physical effect. (Of course, certain deformations will have crossing switches for one observer and none for another. We will take this effect into consideration.) Already in 1926, Kurt Reidemeister Ref. 182 classified those tangle deformations that lead to crossing switches. The classification gives three classes of deformations, today called the three *Reidemeister moves*; they are shown in Figure 43.

- ▷ The ***first Reidemeister move***, or ***type I move***, or ***twist***, is the addition or removal of a twist in a strand.

Reidemeister moves on tangle cores represent the three gauge interactions, whereas the three Reidemeister moves on the vacuum represent (also) gravitational effects.

- ▷ The ***second Reidemeister move***, or ***type II move***, or ***poke***, is the addition or removal of a bend of one strand under (or over) a second strand.
- ▷ The ***third Reidemeister move***, or ***type III move***, or ***slide***, is the displacement of one strand segment under (or over) the crossing of two other strands.

The number of each Reidemeister move is also the number of involved strands. We will discover below that despite appearances, each Reidemeister move induces a crossing switch. To find this connection, we have to generalize the original Reidemeister moves, which were defined in a two-dimensional projection plane, to the three-dimensional case of tangle cores.

Returning to the physics of gauge interactions, for each Reidemeister move we can explore two types of core deformation processes: The first deformation type are *core fluctuations*, which correspond, as we will see, to the emission and absorption of *virtual* interaction bosons. The second type are *externally induced core disturbances*, which correspond to the emission and absorption of *real* interaction bosons. As the first step, we show that in both deformation processes, the first Reidemeister move, the twist, is related to the electromagnetic interaction.

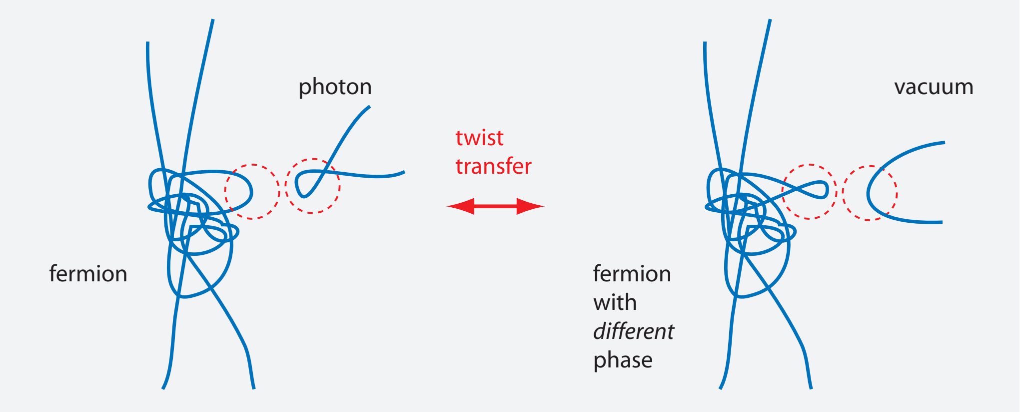


FIGURE 44 A single strand changes the rotation of a tangle: *twist transfer* is the basis of electromagnetism in the strand model. No strand is cut or reglued; the transfer occurs only through the excluded volume due to the impenetrability of strands.

ELECTRODYNAMICS AND THE FIRST REIDEMEISTER MOVE

Experiments show that all interactions are described by potentials. Experiments show that potentials change the rotation frequency and the wave number of wave functions. Interactions result from the absorption and the emission of gauge bosons. In particular, for electromagnetism, the potentials are due to the flow of real and virtual, massless, uncharged spin-1 photons. Photons are emitted from or absorbed by charged particles; neutral particles do not emit or absorb photons. There are two types of charge, positive and negative. The attraction and repulsion of static charges diminishes with the inverse square of the distance. Charge is conserved in nature. All charged particles are massive and move slower than light. The Lagrangian of matter coupled to the electromagnetic field has a U(1) gauge symmetry. There is a single fundamental Feynman diagram. The coupling constant at low energy, the so-called *fine structure constant*, is measured to be 1/137.035 999 1(1); its energy dependence is described by renormalization.

The previous paragraph contains everything known about the electromagnetic interaction. For example, Maxwell's field equations follow from the inverse square law, its relativistic generalization, and the conservation of charge. More precisely, all experimental observations about electricity and magnetism follow from the Lagrangian of quantum electrodynamics, or QED. Therefore we need to show that the Lagrangian of QED follows from the strand model.

STRANDS AND THE TWIST, THE FIRST REIDEMEISTER MOVE

In the strand model of electromagnetism, spin 1 bosons such as the photon are made of a single strand. How can a single strand rotate a tangle? The answer is given in Figure 44: a twisted loop will influence the rotation of a tangle because it changes the possible fluctuations of the tangle core. Due to the impenetrability of strands, an approaching twisted loop will sometimes transfer its twist to the tangle and thus change its phase. The final effect on the phase is the time *average* of all such transfers.

Twisted loops are single strands and can have *two* twist senses. Single strands rep-

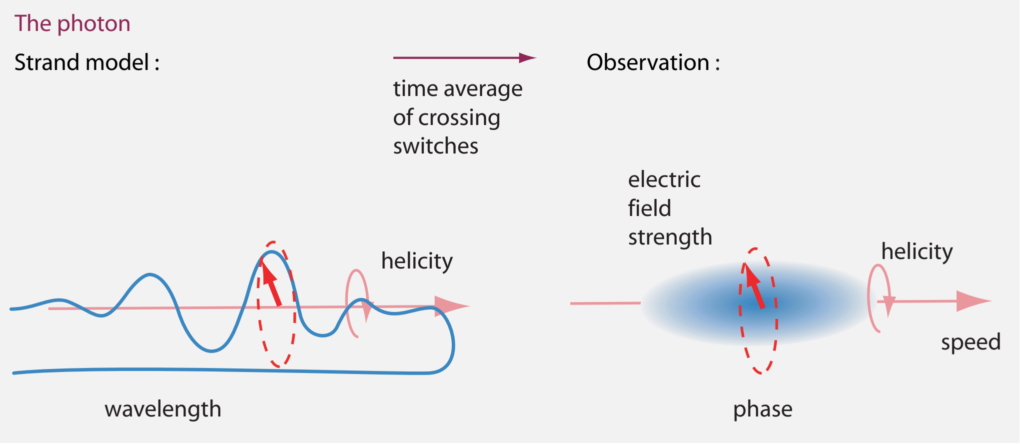


FIGURE 45 The photon in the strand model.

Page 169 resent bosons, as we saw above. Single, twisted but unknotted strands have no mass; in other words, twisted loops move with the speed of light. Twisted loops carry energy. And approaching twisted loops will change the phase,i.e., the orientation of a matter tangle.

It is straightforward to generalize the twist move to arbitrary angles. These general twists can be concatenated. Since they are described by a single angle, they form a U(1) group. We show this is detail shortly. In short, twists behave like *photons* in all their properties. Therefore, the strand model suggests: Page 216

▷ ***Photons*** are single twisted strands. An illustration is given in Figure 45.

▷ The ***electromagnetic interaction*** is the transfer of twists, i.e., the transfer of first Reidemeister moves, between two particles.

The transfer of a twist from a single strand to a tangle core, as shown in Figure 44, thus models the absorption of a photon. We stress again that this transfer results from the way that strands hinder each other's motion, because they are impenetrable. No strand is ever cut or reglued.

Open challenge: Find a better argument for the photon tangle

The argument that leads to the photon tangle is quite hand-waving. Can you give a more precise argument? Challenge 130 ny

Can photons decay or disappear?

The strand model of the photon, as shown in Figure 45, might be seen to suggest that photons can disappear. For example, if a photon strand is straightened out by pulling the ends of the helical deformation, the helix disappears. However, strands have no tension or rigidity, as strands have no observable properties. Thus they cannot be straightened out by pulling. On the other hand, a helix could also disappear by a fluctuation. This is indeed possible, but the energy contained in the helix then is transferred to the strands that generate the fluctuation. Such a process is called the *absorption* of a photon by a charged particle. In other words, energy conservation forbids the decay or disappearance of photons if no charge is involved. Linear and angular momentum conservation also make the same point. Photons are stable particles in the strand model.

ELECTRIC CHARGE

Surrounded by a bath of photon strands, not all fermion tangles will change their rotation state. Only fermion tangles which lack some symmetry will do so. Which symmetry will this be? For a bath of photon strands, thus for a bath that induces Reidemeister I moves, only *chiral* fermion tangles are expected to be influenced. In short:

▷ ***Electric charge*** is due to tangle chirality.*

▷ ***Electrically charged particles*** randomly emit twisted strands. Due to the tangle chirality, a random emission will lead to a slight asymmetry, so that right-handed twists will be in the majority for particles of one charge, and left-handed twists will be in the majority for particles of the opposite charge.

Equating electric charge with tangle chirality allows modelling several important observations. First, because strands are never cut or reglued in the strand model, chirality is a *conserved quantity*. Second, chirality is only possible for tangles that are localized, and thus massive. Therefore, chiral tangles move all slower than light. Third, a static chiral tangle induces a twisted strand density around it that changes as $1/r^2$, as is illustrated in Figure 46. Finally, photons are uncharged: they are not influenced by other photons.

In short, all properties of electric charge found in nature are reproduced by the tangle model. Let us check this in more detail.

CHALLENGE: WHAT KNOT PROPERTY IS ELECTRIC CHARGE?

Mathematicians defined various knot invariants. Several invariants are candidates as building blocks for electric charge: *chirality* c, which can be $+1$ or -1, *minimal crossing number* n, or *topological writhe* w, i.e., the signed minimal crossing number. Page 345

A definition of electric charge q, proposed by Claus Ernst, is $q = c\,(n \bmod 2)$. Another option for the definition of charge is $q = w/3$. We will come back to this issue later on. Page 320

ELECTRIC AND MAGNETIC FIELDS

The definition of photons with twisted strands leads to the following definition.

▷ The ***electric field*** is the oriented volume density of crossings of twisted loops.

▷ The ***magnetic field*** is the flow density of twisted loop crossings.

The simplest way to check this statement is to note that the definition of electric charge as random emission of twisted loops yields Coulomb's expression: the force between two static spherical charges changes with inverse square of the distance. The definition of the

---

Page 322 * The detailed connection between charge and chirality will be explored later on.

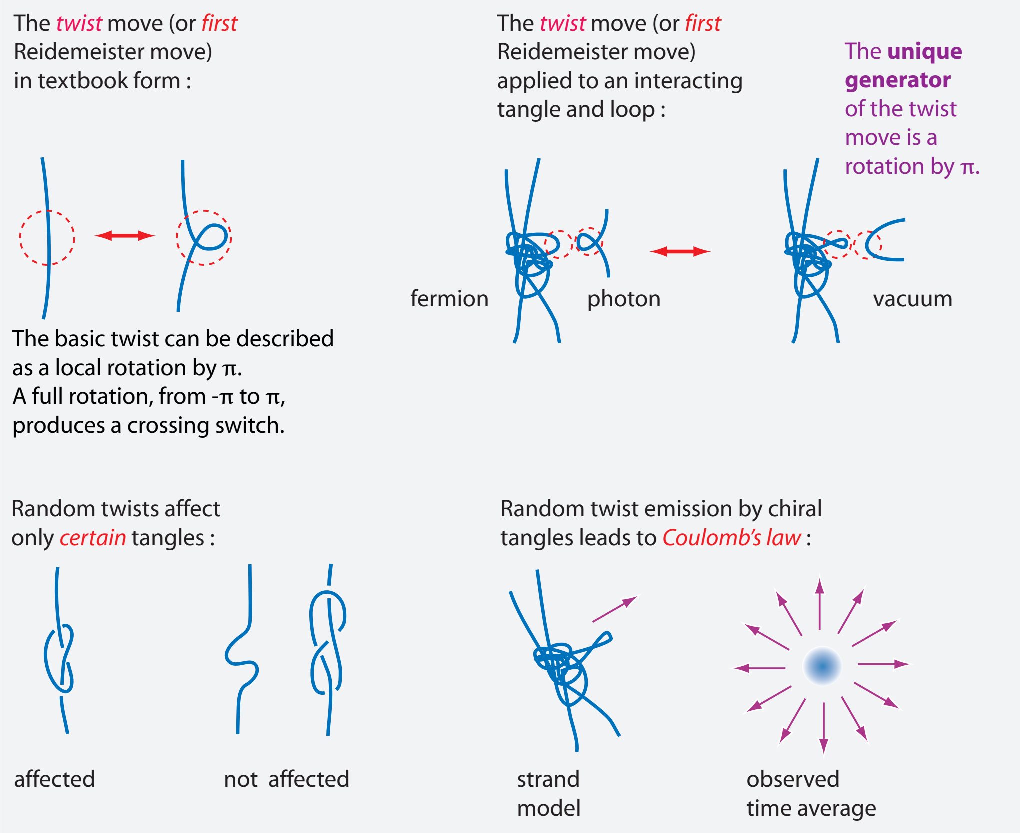


FIGURE 46 Electromagnetism in the strand model.

magnetic field simply follows from that of the electric field by changing to moving frame of reference. The definitions are illustrated in Figure 47.

We note that the electric field is defined almost in the same way as the wave function: Page 173 it is an oriented crossing density. The difference is that the electric field is the crossing density of twisted *loops*, whereas the wave function is the oriented crossing density of *tangles*. The definitions differ only by the topology of the underlying strand structures.

In the strand model, energy is the number of crossing switches per time. The electromagnetic field energy per volume is thus given by the number of crossing switches per volume per time due to twisted loops. Now, the strand model implies, as in the case of Page 180 matter, that *the crossing switch density is proportional to the square of the crossing density*. We have to sum the crossing switches due to the electric and those due to the magnetic field; we also need to multiply each density with a proportionality factor that depends on the units used; in SI units we then get the well-known expression

$$\frac{E}{V} = \frac{\varepsilon_0}{2} E^2 + \frac{1}{2\mu_0} B^2 \,. \qquad (159)$$

The strand model of photons thus reproduces electromagnetic energy.

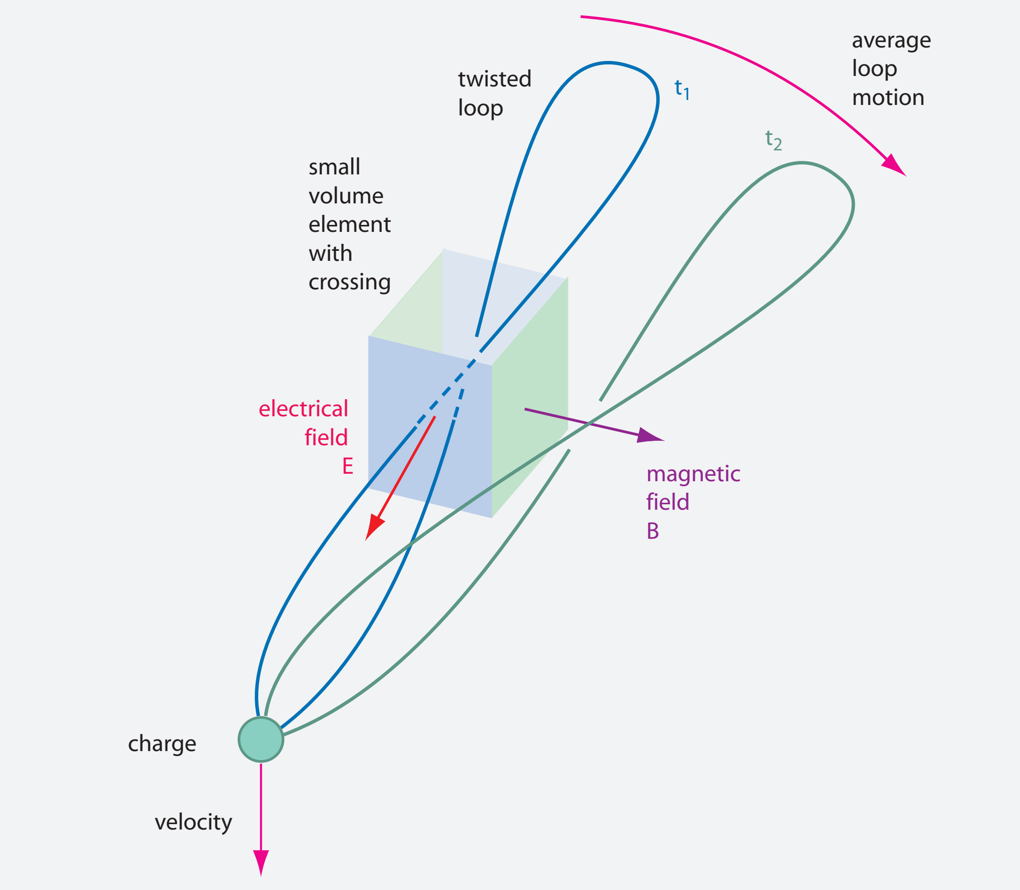


FIGURE 47 Moving strands allow to define electric fields – as the density of loop crossings – and magnetic fields – as the flow of loop crossings.

THE LAGRANGIAN OF THE ELECTROMAGNETIC FIELD

In classical electrodynamics, the energy density of the electromagnetic field is used to deduce its Lagrangian density. The Lagrangian density describes the intrinsic, observer-independent change that occurs in a system. In addition, the Lagrangian density must be quadratic in the fields and be a Lorentz-scalar.

A precise version of these arguments leads to the Lagrangian density of the electromagnetic field

$$\mathcal{L}_{\mathrm{EM}} = \frac{\varepsilon_0}{2} E^2 - \frac{1}{2\mu_0} B^2 = -\frac{1}{4\mu_0} F_{\mu\nu} F^{\mu\nu} \qquad (160)$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \ . \qquad (161)$$

Since the strand model reproduces the electromagnetic energy, it also reproduces the Lagrangian of classical electrodynamics. In particular, Maxwell's equations for the electromagnetic field follow from this Lagrangian density. Maxwell's field equations are thus

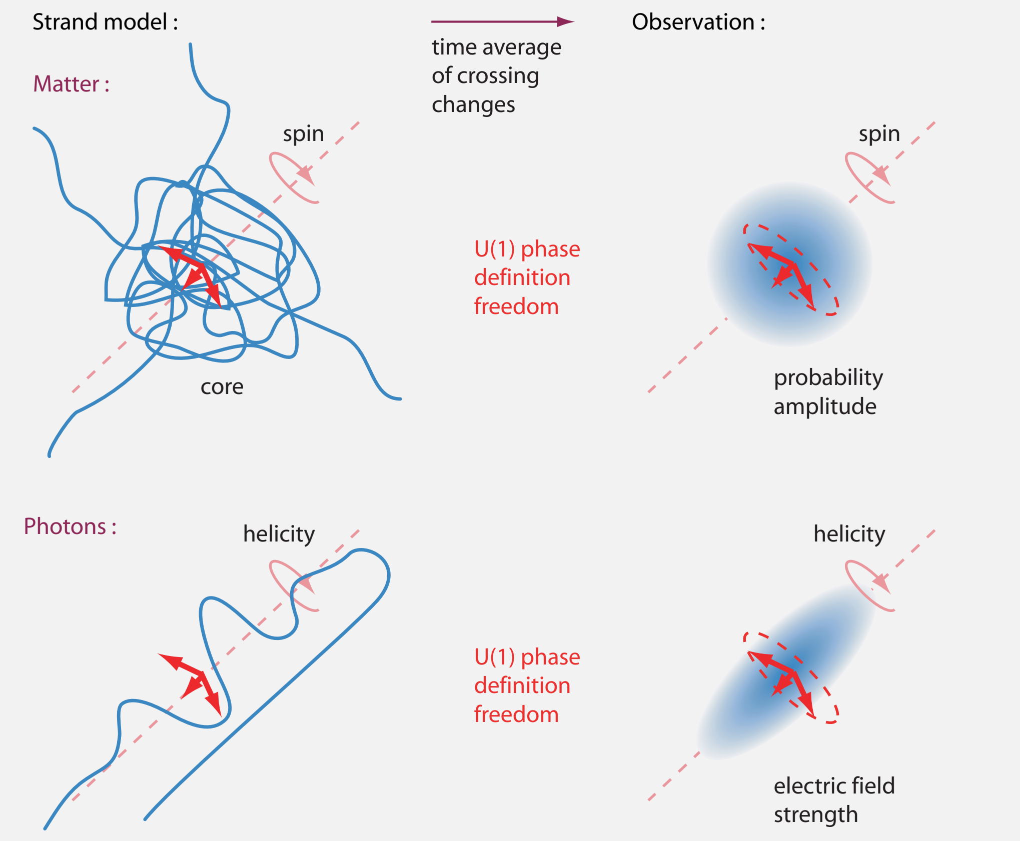


FIGURE 48 The choice of gauge in the case of electrodynamics.

a consequence of the strand model. Obviously, this is no news, because any model that reproduces Coulomb's inverse square distance relation and leaves the speed of light invariant automatically contains Maxwell's field equations.

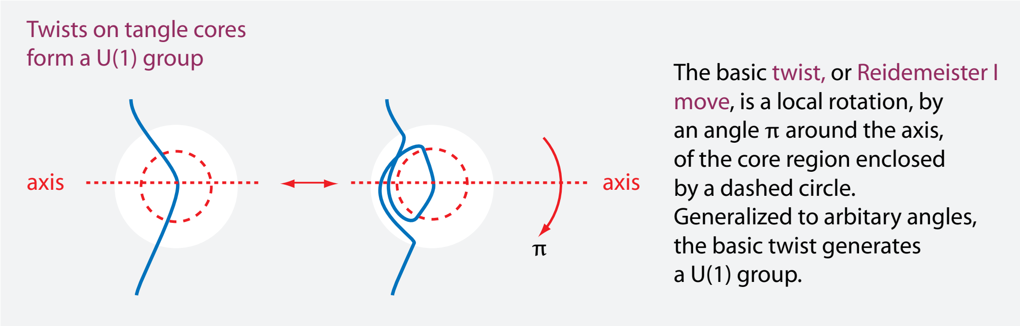
U(1) GAUGE INVARIANCE INDUCED BY TWISTS

In nature, the potential A_μ is not uniquely defined: there is a freedom in the choice of gauge. The change from one gauge to another is a *gauge transformation*. Gauge transformations are thus transformations of the potentials that have no effect on observations. In particular, gauge transformations leave unchanged all field intensities and field energies on one hand and particle probabilities and particle energies on the other hand.

In the strand model the following definitions are natural:

- ▷ A ***gauge choice*** for *radiation* and for *matter* is the choice of definition of the respective phase arrow.
- ▷ A ***gauge transformation*** is a change of definition of the phase arrow.

Various gauge choices for electrodynamic systems are illustrated in Figure 48. We see directly that the range of choices of the phase arrow forms a circle around the spin axis.



The basic twist, or Reidemeister I move, is a local rotation, by an angle π around the axis, of the core region enclosed by a dashed circle. Generalized to arbitary angles, the basic twist generates a U(1) group.

FIGURE 49 How the set of generalized twists – the set of all local rotations of a single strand segment around an axis – forms a U(1) group group.

We note that gauge transformations have no effect on the flow of crossing switches. In other words, gauge transformations leave electromagnetic field intensities and thus also the energy of the electromagnetic field invariant, as observed. Similarly, gauge transformations have no effect on the number of crossing switches of rotating tangles. A rotation by 4π remains the same, independently of which definition of arrow is chosen. Therefore, gauge transformations leave probability densities and phase differences unchanged, as observed.

From the gauge choice there is only a small step to the gauge interaction. We recall:

▷ A ***gauge interaction*** is a change of phase by strand deformation in a particle core.

Electromagnetism results from the transfer of twists. The basic twist, or first Reidemeister move, corresponds to a local rotation of some strand segment in the core by an angle π, as illustrated by Figure 49. The twists of highest importance are those whose rotation axis is along the direction of the spin axis. To make the discussion as specific as possible, but without loss of generality, we choose this rotation axis in the following.

Twists can be generalized to arbitrary angles: we simply define a *generalized twist* as a local rotation of a strand segment by an arbitrary angle. Generalized twists can be concatenated, and the identity twist – no local rotation at all – also exists. Generalized twists thus form a group. Figure 48 makes clear that twists rotate the phase arrow around its axis, i.e., along a circle. In particular, a generalized twist that rotates the tangle phase of a fermion or the strand phase of a photon by an angle 2π (the arrow itself by 4π) is equivalent to the identity twist. These properties uniquely define the group U(1). In short, Figure 49 shows that twists define the group U(1), which has the topology of a circle.

In summary, the addition of a twist to a fermion tangle or to a photon strand changes their phase, and thus represents a gauge interaction. We have shown that core fluctuations induced by twists produce a U(1) gauge symmetry. Electromagnetic field energy and particle energy are U(1) invariant. In short, the strand model implies that *the gauge group of quantum electrodynamics is U(1).* With this result, we are now able to deduce the Lagrangian of QED.

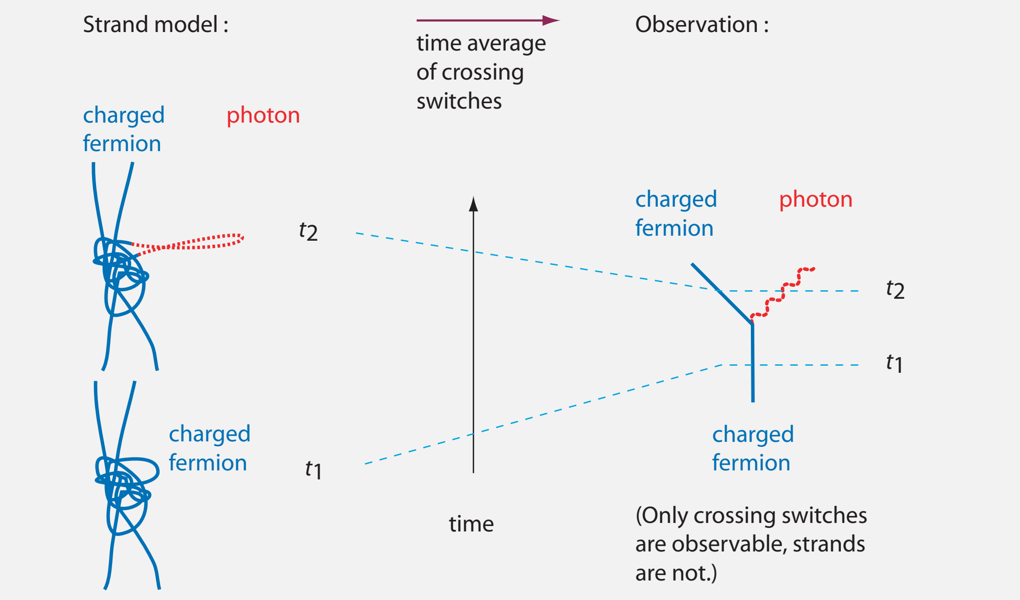


FIGURE 50 The fundamental Feynman diagram of QED and its tangle version.

THE LAGRANGIAN OF QED

Given the U(1) gauge invariance of observables, the Lagrangian of quantum electrodynamics, QED, follows directly, because U(1) gauge invariance is equivalent to minimal coupling. The Lagrangian density $\mathcal{L}$ of a *neutral*, *free*, and relativistic fermion in an electromagnetic field is

$$\mathcal{L} = \overline{\Psi}(i\hbar c \not\partial - mc^2)\Psi - \frac{1}{4\mu_0} F_{\mu\nu} F^{\mu\nu} . \qquad (162)$$

Page 205

We deduced the fermion term above, when discussing quantum theory, and the electromagnetic term just now, from the properties of twisted loops. Minimal coupling changes this Lagrangian density into the Dirac Lagrangian density of a *charged*, i.e., *interacting*, relativistic fermion in the electromagnetic field, in other words, into the Lagrangian density of QED:

$$\mathcal{L}_{\text{QED}} = \overline{\Psi}(i\hbar c \not{D} - mc^2)\Psi - \frac{1}{4\mu_0} F_{\mu\nu} F^{\mu\nu} . \qquad (163)$$

Here, $\not{D} = \gamma^\sigma \mathrm{D}_\sigma$ is the *gauge covariant derivative* that is defined through minimal coupling to the charge q:

$$\mathrm{D}_\sigma = \partial_\sigma - iqA_\sigma . \qquad (164)$$

The Lagrangian density of QED is indeed invariant under U(1) gauge transformations. We have thus recovered the Lagrangian density of quantum electrodynamics from strands. Strands thus reproduce the most precisely tested theory of physics.

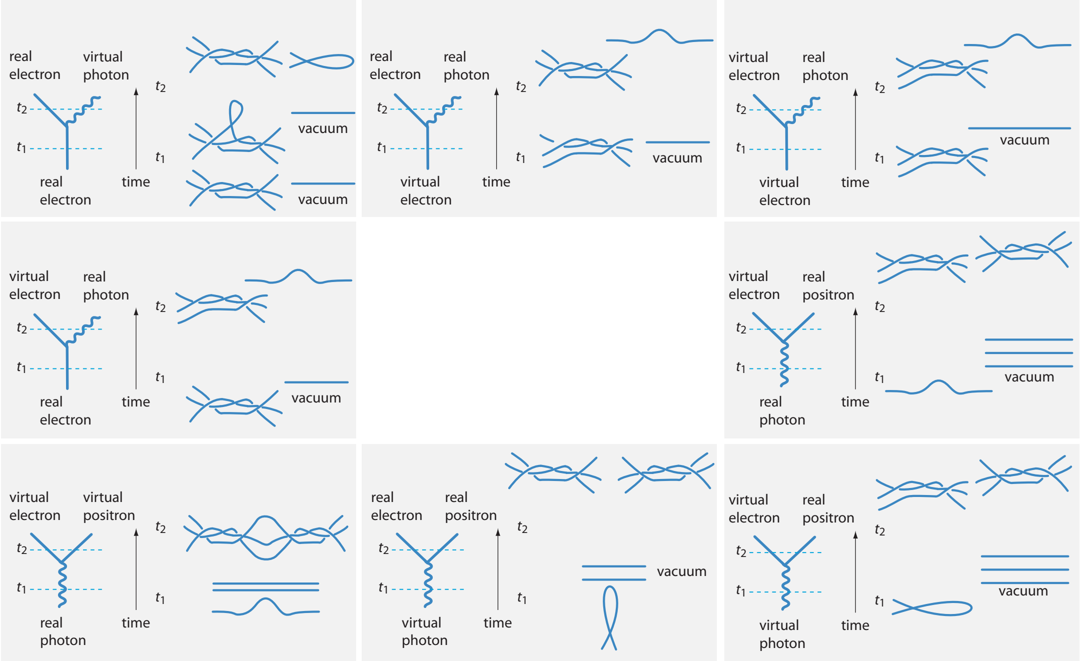


FIGURE 51 The different options of the fundamental Feynman diagram of QED and their tangle versions.

FEYNMAN DIAGRAMS AND RENORMALIZATION

Feynman diagrams are abbreviations of formulas to calculate effects of quantum electrodynamics in perturbation expansion. Feynman diagrams follow from the Lagrangian of QED. All Feynman diagrams of QED can be constructed from one fundamental diagram, shown on the right-hand side of Figure 50.

In the strand model, the fundamental Feynman diagram can be visualized directly in terms of strands, as shown on the left-hand side of Figure 50. This is the same diagram that we have explored right at the start of the section on electrodynamics, when we defined electrodynamics as twist exchange. (The precise tangles for the charged fermions will be deduced later on.) Since all possible Feynman diagrams are constructed from the fundamental diagram, the strand model allows us to interpret all possible Feynman diagrams as strand diagrams. For example, the strand model implies that the vacuum is full of virtual particle-antiparticle pairs, as shown in Figure 52.

Page 211

Page 273

In quantum field theory, Lagrangians must not only be Lorentz and gauge invariant, but must also renormalizable. The strand model makes several statements on this issue. At this point, we focus on QED only; the other gauge interactions will be treated below. The strand model reproduces the QED Lagrangian, which is renormalizable. Renormalizability is a natural consequence of the strand model (assuming that strand diameters are negligible). The reason for renormalizability that the strand model reproduces the single, fundamental Feynman diagram of QED, without allowing other types of diagrams.

The twist deformations underlying the strand model for QED also suggest new ways to calculate higher order Feynman diagrams. Such ways are useful in calculations of

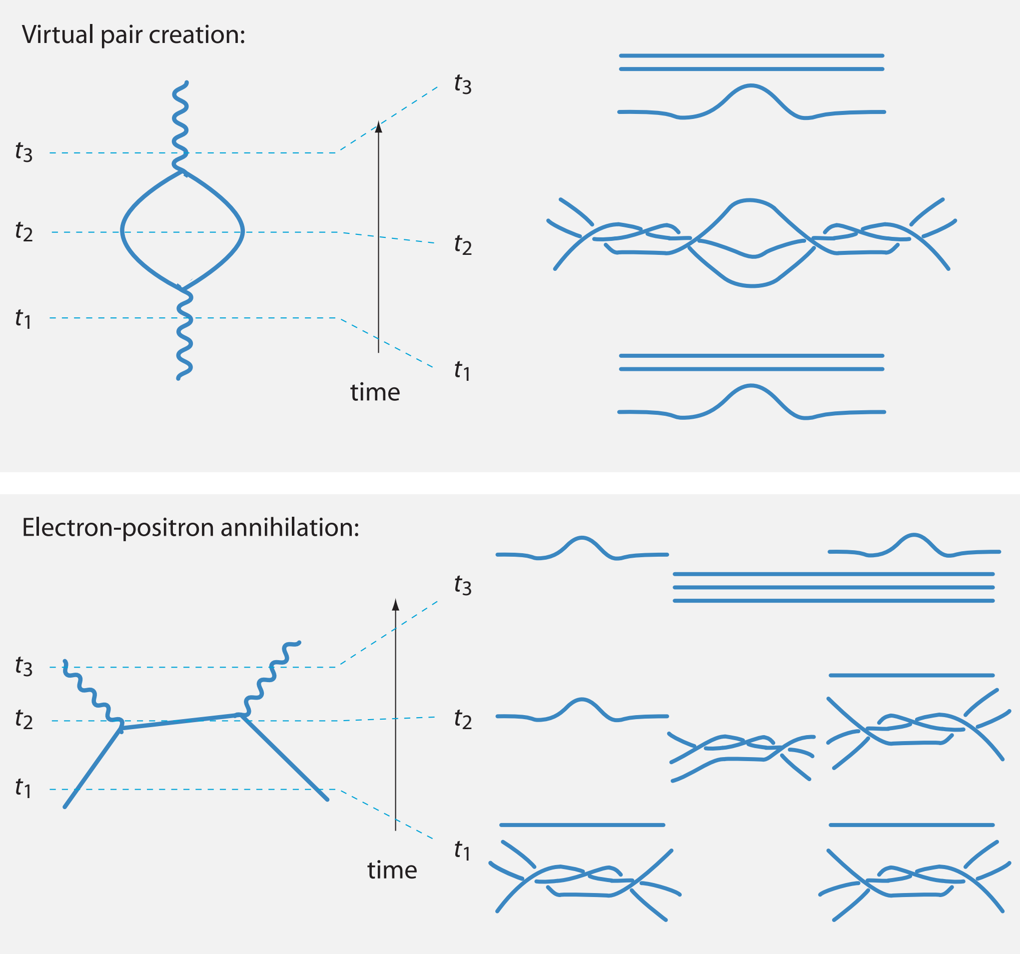


FIGURE 52 Some Feynman diagram of QED with their tangle versions.

g-factors of charged particles. In particular, the strand model for QED, as shown in Figure 50, implies that higher order QED diagrams are simple *deformations* of lower order diagrams. Taking statistical averages of strand deformations up to a given number of crossings thus allows to calculate QED effects up to a given order in the coupling. However, using this approach for calculations is not a topic of the present text.

The strand model also suggests that the difference between renormalized and unrenormalized mass and charge reflects the difference between minimal and non-minimal crossing switch number, or equivalently, between simple and more complex tangle deformations, where strands are deformed on smaller distance scales. In other terms, unrenormalized quantities (at Planck energy) can be imagined as those deduced when the tangles are pulled tight, i.e., pulled to Planck distances, whereas renormalized mass and charge values are those deduced for particles surrounded by many large-size fluctuations.

In summary, the strand model provides a new underlying picture or mechanism for Feynman diagrams. The strand model does not change any physical result at any experimentally accessible energy scale. In particular, the measured running with energy of the fine structure constant and of the masses of charged particles are reproduced by the strand model, because Feynman diagrams of all orders are reproduced. Deviations are only expected near the Planck energy, when tangles are pulled tight.

FUN CHALLENGES ABOUT QED

In the strand model, it should be possible to deduce that the anomalous magnetic moment of the electron is given by

$$\frac{g}{2} = 1 + \frac{\alpha}{2\pi} - O(\alpha^2) \ . \tag{165}$$

Challenge 131 d Find a simple explanation.

MAXWELL'S EQUATIONS

The strand model also allows to check Maxwell's field equations of electrodynamics directly. The equations are:

$$\begin{aligned} \nabla\,\mathbf{E} &= \frac{\rho}{\varepsilon_0} \ , \\ \nabla\,\mathbf{B} &= 0 \ , \\ \nabla\times\mathbf{E} &= -\frac{\partial\mathbf{B}}{\partial t} \ , \\ \nabla\times\mathbf{B} &= \frac{1}{c^2}\frac{\partial\mathbf{E}}{\partial t} + \mu_0\mathbf{J} \quad . \end{aligned} \tag{166}$$

- The first of these equations is satisfied whatever the precise mechanism at the basis of loop emission by electric charges may be. Indeed, any mechanism in which a charge randomly sends out or swallows a handle yields a $1/r^2$ dependence for the electrostatic field and the required connection between charge and the divergence of the electric field. This is not a deep result: any spherically-symmetric system that randomly emits or swallows some entity produces the equation, including the underlying inverse-square dependence. The result can also be confirmed in another, well-known way. In any exchange interaction between two charges, the exchange time is proportional to their distance apart r; in addition, quantum theory states that the exchanged momentum is inversely proportional to the distance r. Therefore, the force, or momentum per unit time, varies as $1/r^2$. The reason that this relation is valid independently of the underlying motion of the loops is due to the three-dimensionality of space: all localized sources automatically fulfil the inverse square dependence.

The constant on the right hand side of the first equation results from the definition of the units; in the language of the strand model, the constant fixes the loop emission rate for an elementary charge.

- The second of the field equations (166) expresses the lack of magnetic charges. This equation is automatically fulfilled by the strand model, as the definition of the magnetic field with strands does not admit any magnetic sources. In fact, strands suggest that no localized entity can have a magnetic charge. Also this equation is valid independently of the details of the motion of the strands. Again, this is a topological effect.

- The third field equation relates the temporal change of the magnetic field to the curl of the electric field. In the strand model, this is satisfied naturally, because a curl in the electric field implies, by construction, a change of the magnetic field, as shown by Figure 47. Again, this relation is valid independently of the details of the motion of the strands, as long as the averaging scale is taken to be large enough to allow the definition of electric and the magnetic fields.
- The most interesting equation is the last of the four Maxwell equations (166): in particular, the second term on the right-hand side, the dependence on the charge current. In the description of electrodynamics, the charge current **J** appears with a positive sign and with no numerical factor. (This is in contrast to linearized gravity, where the current has a numerical factor and a negative sign.) The positive sign means that a larger current produces a larger magnetic field. The strand model reproduces this factor: strands lead to an effect that is proportional both to charge (because more elementary charges produce more crossing flows) and to speed of movement of charge (large charge speed lead to larger flows). Because of this result, the classical photon spin, which is defined as L/ω, and which determines the numerical factor, namely 1, that appears before the charge current **J**, is recovered. Also this connection is obviously independent of the precise motion of the underlying strands.

The first term on the right-hand side of the fourth equation, representing the connection between a changing electric field and the curl of the magnetic field, is automatically in agreement with the model. This can again be checked from Figure 47 – and again, this is a topological effect, valid for any underlying strand fluctuation. As an example, when a capacitor is charged, a compass needle between the plates is deflected. In the strand model, the accumulating charges on the plates lead to a magnetic field. The last of Maxwell's equations is thus also confirmed by the strand model.

In summary, the strand model reproduces Maxwell's equations. However, this is not a great feat. Maxwell-like equations appear in many places in field theory, for example in solid-state physics and hydrodynamics. Mathematical physicists are so used to the appearance of Maxwell-like equations in other domains that they seldom pay it much attention. The real test for any model of electrodynamics is the deviation that it predicts from electrodynamics.

SUMMARY ON QED AND EXPERIMENTAL PREDICTIONS

In the strand model, photons are single, helically twisted strands, randomly exchanged between charges; charges are chiral tangles. This is the complete description of QED using strands.

We have shown that twists of tangle cores lead to U(1) gauge invariance. Maxwell's equations of electrodynamics follow from the exchange of twists. In short, we have deduced all experimental properties of quantum electrodynamics, except one: the strength of the coupling. Despite this open point, we have settled one line of the millennium list of open issues in fundamental physics. Page 158

Is there a difference between the strand model and quantum electrodynamics? The answer is subtle. There are *no measurable* differences between the strand model and QED. For example, the g-factor of the electron or the muon predicted by QED is not changed by the strand model. The U(1) gauge symmetry and the whole of QED remain valid at all energies. There are no other gauge groups. QED remains exact in all cases – as long as gravity plays no role.

This prediction of the strand model is disconcerting. There is thus *no* grand unification in nature; there is no general gauge group in nature, be it SU(5), SO(10), E6, E7, E8, SO(32) or any other. This fact indirectly also rules out supersymmetry and supergravity. This conclusion contrasts with many cherished habits of thought.

The strand model thus widens the validity of QED and puts the spot on its limitation. The equivalence of Feynman diagrams and strand diagrams implies that deviations of the strand model from QED are expected *only* when gravity starts to play a role. The strand model predicts that this will only happen near the Planck energy $\sqrt{\hbar c^5/4G}$. At lower energies, QED is predicted to remain valid.

The strand model also confirms that the combination of gravity and quantum theory turns all Planck units into *limit* values, because there is a maximum density of strand crossings in nature, due to the fundamental principle. For example, the maximum speed is c, the maximum elementary particle energy is the Planck energy, and the shortest measurable length is the Planck length $\sqrt{4G\hbar/c^3}$. In the same way, the strand model predicts a maximum electric field value $E_{\text{max}} = c^4/4Ge \approx 1.9 \cdot 10^{62}$ V/m and a maximum magnetic field value $B_{\text{max}} = c^3/4Ge \approx 6.3 \cdot 10^{53}$ T. All physical systems – including all astrophysical objects, such as gamma ray bursters or quasars – are predicted to conform to this limit. So far, this prediction agrees with observations.

All limit values for observables have a simple explanation: limit values appear when strands are as closely packed as possible. In the strand model, strands cannot be packed more closely than to Planck distances. Thus the strand model predicts that approaching the electric or magnetic field limit values – given by quantum gravity – is the only option to observe deviations from QED. This is most probably impossible, so that we can state that there are no measurable difference between the strand model and QED.

In short, our exploration of QED has left open only two points: the calculation of the electromagnetic coupling constant and the determination of the precise tangle for each elementary particle. Before we clarify these points, we look at the next Reidemeister move.

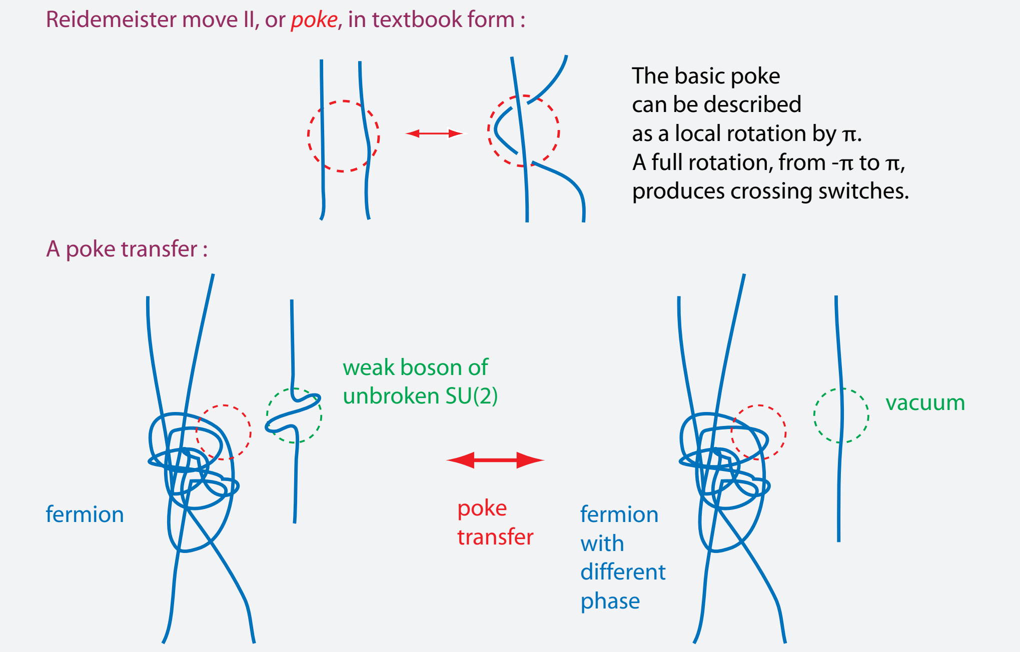
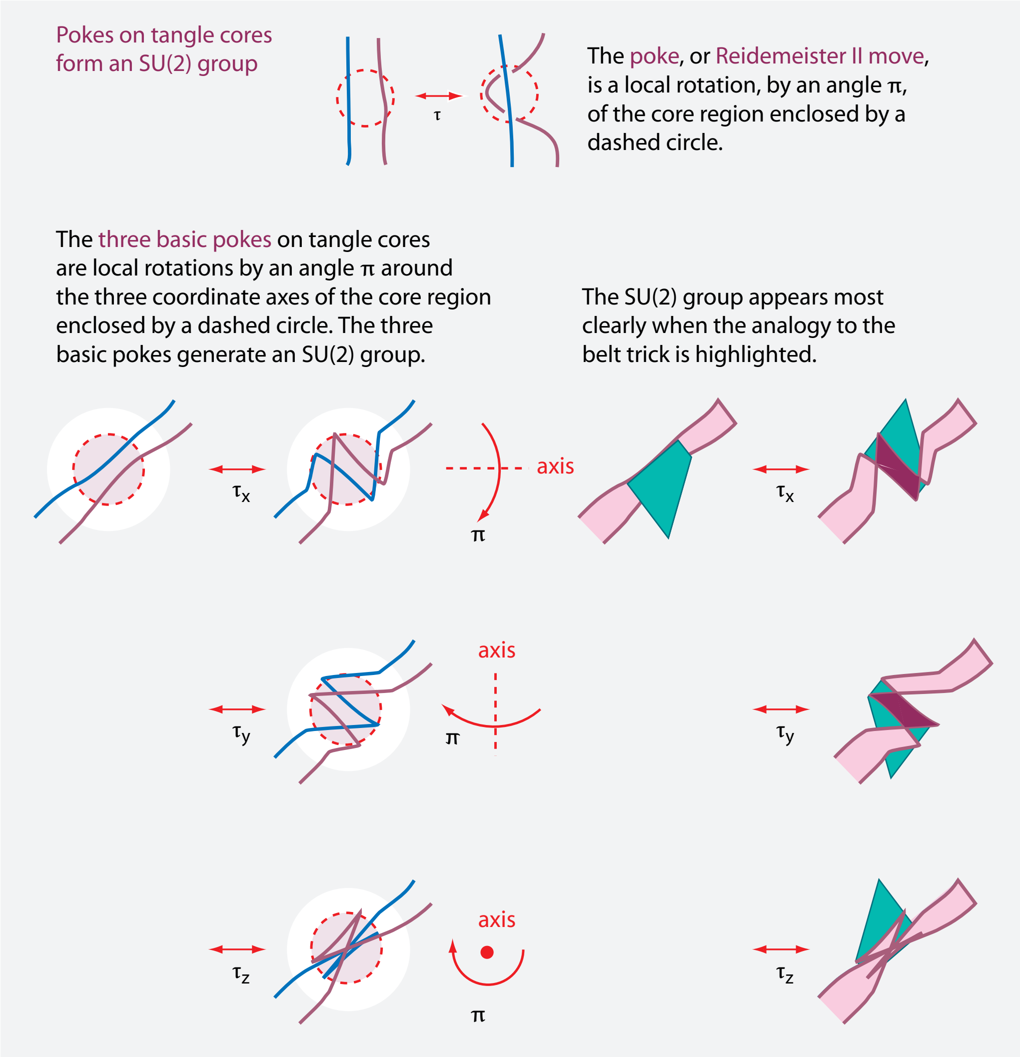


FIGURE 53 *Poke transfer* is the basis of the weak interaction in the strand model. No strand is cut or reglued; the transfer occurs only through the excluded volume due to the impenetrability of strands.

THE WEAK NUCLEAR INTERACTION AND THE SECOND REIDEMEISTER MOVE

In nature, the weak interaction is the result of the absorption and the emission of massive spin-1 bosons that form a broken weak triplet. The W and the Z bosons are emitted or absorbed by particles with weak charge; these are the left-handed fermions and right-handed antifermions. In other words, the weak interaction breaks parity P maximally. The W boson has unit electric charge, the Z boson has vanishing electric charge. The emission or absorption of W bosons changes the particle type of the involved fermion. The weak bosons also interact among themselves. All weakly charged particles are massive and move slower than light. The Lagrangian of matter coupled to the weak field has a broken SU(2) gauge symmetry. There are a few fundamental Feynman diagrams with triple and quartic vertices. The weak coupling constant is determined by the electromagnetic coupling constant and the weak boson masses; its energy dependence is fixed by renormalization.

The previous paragraph summarizes the main observations about the weak interaction. All observations related to the weak interaction are described and contained in its Lagrangian. Therefore, we need to show that the weak interaction Lagrangian follows from the strand model.


Pokes on tangle cores form an SU(2) group

The poke, or Reidemeister II move, is a local rotation, by an angle π, of the core region enclosed by a dashed circle.

The three basic pokes on tangle cores are local rotations by an angle π around the three coordinate axes of the core region enclosed by a dashed circle. The three basic pokes generate an SU(2) group.

The SU(2) group appears most clearly when the analogy to the belt trick is highlighted.

FIGURE 54 How the set of all pokes – the set of all deformations induced on tangle cores by the weak interaction – forms an SU(2) gauge group. The relation to the belt trick, with a pointed buckle and two belts, is also shown.

Strands, pokes and SU(2)

Page 207

As explained above, any gauge interaction involving a fermion is a deformation of the tangle core that changes the phase and rotation of the fermion tangle. We start directly with the main concepts for the weak interaction.

- ▷ The ***weak interaction*** is the transfer of a poke, i.e., the transfer of a Reidemeister II move, between two particles. An illustration is given in Figure 53. Strands are not cut in this process; they simply transfer the deformation as a result of their impenetrability.
- ▷ ***Weak intermediate bosons*** are families of knotted single strands. An illus-

tration is given in Figure 56.

Strands describe the weak interaction as exchange of pokes. In tangle cores, the *basic* pokes induce local rotations by an angle π, as shown on the top of Figure 53: each basic poke rotates the region enclosed by the dotted circle. Given a spin axis, there are *three*, linearly independent, basic pokes, in three mutually orthogonal directions. The three basic pokes τ_x, τ_y and τ_z are illustrated in Figure 54. This figure also shows that the three pokes act on the local region in the same way as the three possible mutually orthogonal rotations act on a belt buckle.

In particular, the pokes of Figure 54 show that the product of two different basic pokes gives the third poke, together with a sign – which depends on whether the sequence is cyclic or not – and a factor of i. Using the definition of -1 as a local rotation by 2π, we also find that the square of each basic poke is -1. In detail, we can read off the following multiplication table for the three basic pokes:

Challenge 132 e

Page 179

$$\begin{array}{c|ccc} \cdot & \tau_x & \tau_y & \tau_z \\ \hline \tau_x & -1 & i\tau_z & -i\tau_y \\ \tau_y & -i\tau_z & -1 & i\tau_x \\ \tau_z & i\tau_y & -i\tau_x & -1 \end{array} \qquad (167)$$

In other terms, the three basic pokes form the generators of an SU(2) group. When seen as local rotations, pokes can be generalized to arbitrary angles, and they can be concatenated. We thus find that general pokes form the full SU(2) group. We already expected this from the equivalence with the belt trick.

In summary, we can state that in any definition of the phase of a tangled fermion core, there is an SU(2) gauge freedom. In other words, *the gauge group of the unbroken weak interaction is SU(2).*

Weak charge and parity violation

Surrounded by a bath of strands that continuously induce pokes for a long time, not all tangles will change their phase. Only tangles that lack some symmetry will do so. One symmetry that must be lacking is spherical symmetry. Therefore, only tangles whose cores lack *spherical symmetry* have the chance to be influenced by random pokes. Since all tangles with cores lack spherical symmetry, all such tangles, i.e., all massive particles, are candidates to be influenced, and thus are candidates for weakly charged particles. We therefore explore them in detail.

If a tangle is made of a *single* knotted strand, we expect it to be influenced by large numbers of pokes. Such tangle cores are massive spin-1 bosons; they are rotating continuously, and the rotation will induce a left-right asymmetry that will lead to a higher effect of a poke than of its reverse. Single knotted strands are thus predicted to carry weak charge. We therefore expect that the weak bosons themselves interact weakly. In other words, the strand model predicts that the weak interaction is a *non-Abelian* gauge theory.

In contrast, if a tangle is made of a single *unknotted* strand, it is not affected by random pokes. The strand model thus predicts that the photon has no weak charge, as is observed.

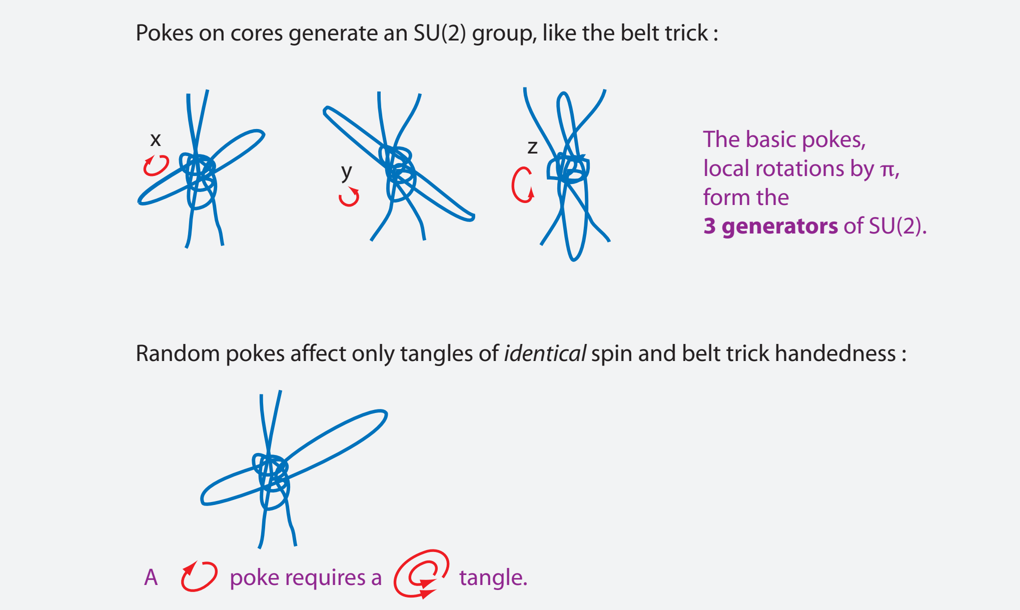


FIGURE 55 The three basic pokes and weak charge in the strand model.

(The same also holds for gluons, as we will see.)

If a tangle is made of *two or more* strands, it represents a massive spin-1/2 particle. All such fermion cores lack spherical symmetry. When a fermion spins, two things happen: the core rotates and the belt trick occurs, untangling the tails. Both the rotation and the untangling can be either left-handed or right-handed, giving four combinations. (As explained above, these combinations form the four components of the Dirac spinor.) Now, in order to feel any average effect when large numbers of random pokes are applied, a core must undergo different effects for a poke and its reverse. This will only happen if the core rotation and the tail untangling are of the same handedness. The strand model thus predicts that random pokes will only affect a core if the core rotation and the belt trick are of the *same* handedness. In other words, random pokes will only affect left-handed particles or right-handed antiparticles.

Page 200

▷ Non-vanishing ***weak charge*** for fermions appears only for tangle cores whose spin and untangling handedness match.

Thus, the strand model predicts that the weak interaction violates parity *maximally*, as is observed. In other terms, the weak charge and the parity violation of the weak interaction are related to the belt trick.

The strand definition of weak charge leads to two conclusions that can be checked by experiment. First, the belt trick is only possible for tangles that are localized, and thus massive. In short, only massive particles interact weakly. In fact, *all* massive particles interact weakly. In other words, all weakly charged particles are predicted to move more slowly than light. Secondly, all electrically charged particles are predicted to be also weakly charged. Both conclusions agree with observation.

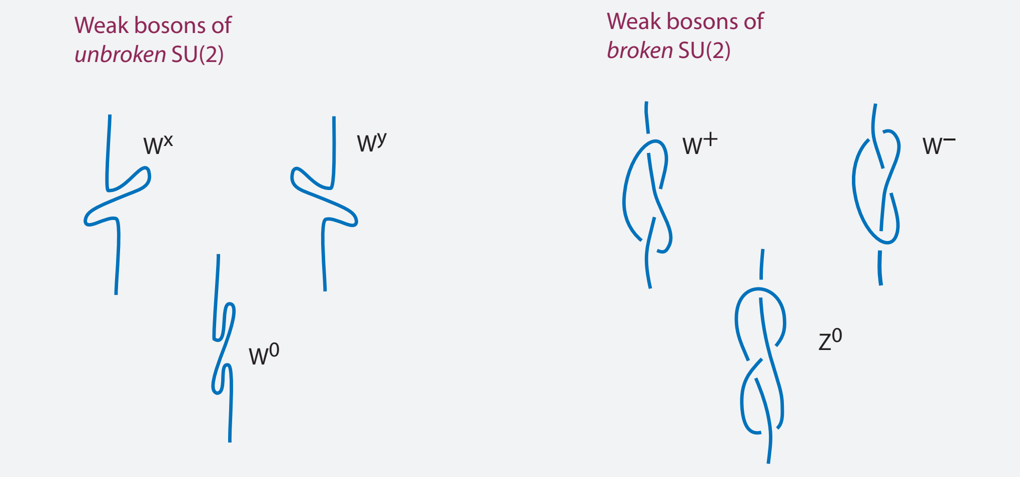


FIGURE 56 Poke-inducing strands (left) differ from weak vector bosons (right) because of symmetry breaking. The figure shows only the simplest possible tangles for each weak gauge boson.

In short, all properties of weak charge found in nature are reproduced by the tangle model. In particular, core fluctuations induced by pokes produce a SU(2) gauge symmetry.

Weak bosons

Gauge bosons are those particles that are exchanged between interacting fermions; gauge bosons induce phase changes of fermions. In the strand model, all spin-1 bosons are made of a single strand. Single strands that induce phase changes in fermions interacting weakly are shown on the left side of Figure 56. They generate the three basic pokes τ_x, τ_y and τ_z. Unknotted strands, however, are *massless*. In the strand model, single strands that induce pokes *differ* from the knotted weak intermediate bosons, as shown on the right of Figure 56. This difference is due to the breaking of the SU(2) gauge symmetry, as we will find out soon.

The energy of the weak field is given by the density of weak gauge boson strands. As long as the SU(2) symmetry is not broken, the energy of the weak field and the energy of fermions are both SU(2) invariant. As a consequence, we are now able to deduce a large part of the Lagrangian of the weak interaction, in the case that the SU(2) symmetry is unbroken.

The Lagrangian of the unbroken SU(2) gauge interaction

As long as SU(2) is unbroken, the vector bosons are described as unknotted strands that induce pokes, as shown in Figure 56. There are three such bosons. Since they are made of a single strand, they have spin 1; since they are unknotted, they have zero mass and electric charge.

Energy is the number of crossing switches per time. As long as SU(2) is unbroken and the weak bosons are massless, the energy of the weak boson field und thus their Lagrangian density is given by the same expression as the energy of the photon field. In

Page 215

particular, the energy density is quadratic in the field intensities. We only have to add the energies of all three bosons together to get:

$$\mathcal{L} = -\frac{1}{4} \sum_{a=1}^{3} W^{a}{}_{\mu\nu} W_{a}{}^{\mu\nu} \,, \tag{168}$$

This expression is SU(2) gauge invariant. Indeed, SU(2) gauge transformations have no effect on the number of crossing switches due to weak bosons or to the motion of pokes. Thus, gauge transformations leave weak field intensities and thus also the energy of the weak fields invariant, as observed.

We can now write down the Lagrangian for weakly charged fermions interacting with the weak vector bosons. Starting from the idea that tangle core deformations lead to phase redefinitions, we have found that pokes imply that the *unbroken* weak Lagrangian density for matter and radiation fields is SU(2) gauge invariant. Using minimal coupling, we thus get the Lagrangian

$$\mathcal{L}_{\text{unbroken weak}} = \sum_{f} \overline{\Psi}_{f} (i\hbar c \not{D} - m_{f} c^{2}) \Psi_{f} - \frac{1}{4} \sum_{a=1}^{3} W^{a}{}_{\mu\nu} W_{a}{}^{\mu\nu} \,, \tag{169}$$

where $\not{D}$ is now the SU(2) gauge covariant derivative and the first sum is taken over all fermions. In this Lagrangian, only the left-handed fermions and the right-handed antifermions carry weak charge. This Lagrangian, however, does not describe nature: the observed SU(2) breaking is missing.

SU(2) BREAKING

In nature, the weak interaction does *not* have an SU(2) gauge symmetry. The symmetry is said to be *broken*. The main effect of SU(2) symmetry breaking are the non-vanishing masses of the W and Z bosons, and thus the weakness and the short range of the weak interaction. In addition, the symmetry breaking implies electroweak unification.

The strand model suggests the following description:

▷ ***Mass generation*** for bosons and the related SU(2) ***symmetry breaking*** are due to *overcrossing* at the border of space. Figure 57 illustrates the idea.

In this description, overcrossing is assumed to occur at the border of space, more precisely, in a region where physical space is not defined any more; in such a region, overcrossing is *not* forbidden and *can* occur. The probability of overcrossing is low, because the crossings have first to fluctuate to that region and then fluctuate back. Nevertheless, the process can take place. Overcrossing appears *only* in the weak interaction. It does not appear in the other two gauge interactions, as the other Reidemeister moves do not allow processes at the border of space. In the strand model, this is the reason that only SU(2) is broken in nature. In short, SU(2) breaking is a natural consequence of the second Reidemeister move.

Overcrossing transforms the unknotted, and thus massless poke strands into the knot-

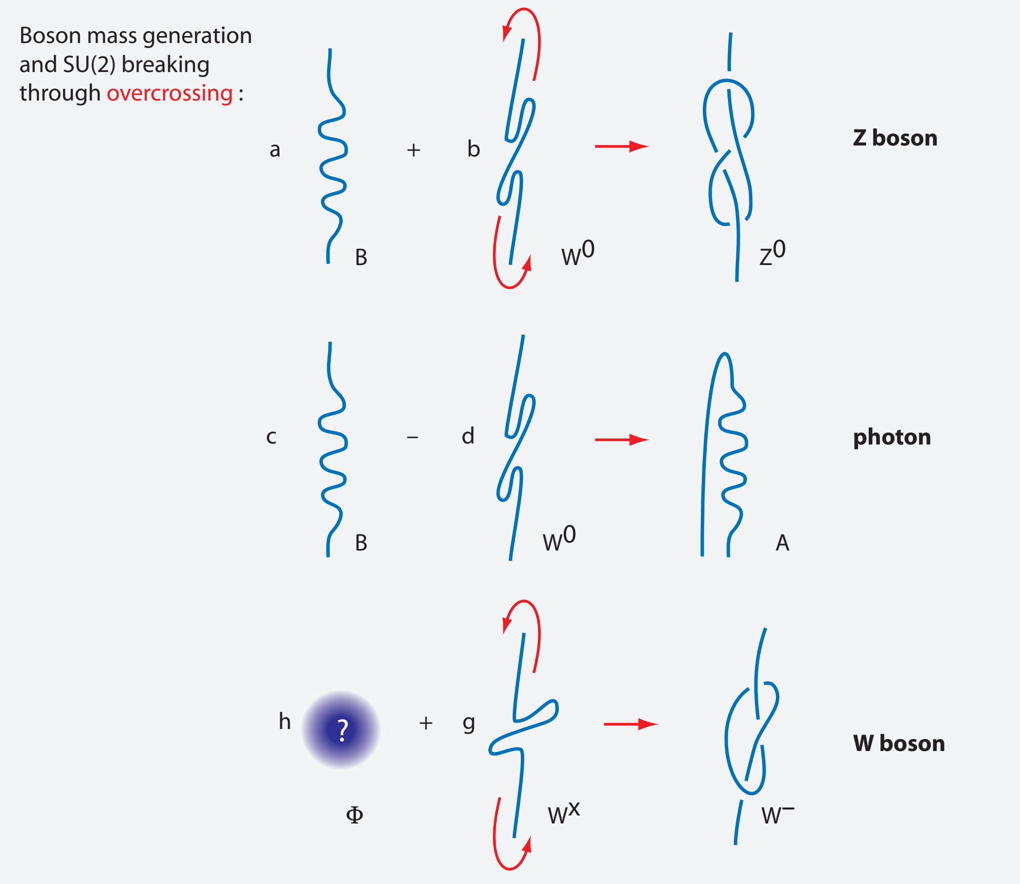


FIGURE 57 In the strand model, mass generation and the breaking of the SU(2) gauge symmetry in the weak interaction is due to overcrossing (a, b, c, d, h and g are scalar factors); the third process, involving the Higgs boson, will be clarified later on.

ted, and thus massive W and Z strands. Overcrossing is thus a mass-generating process. The precise mass values that it generates will be determined below. The strand model thus confirms that mass generation is related to the breaking of the weak interaction. Page 306

When producing the mass of the Z boson, overcrossing mixes it with the 'original' photon. This is shown in Figure 57. The mixing is due to their topological similarities in the strand model. The Z boson is achiral, and thus electrically neutral, as observed. The existence of a neutral, massive Z boson implies that elastic neutrino scattering in matter occurs in nature, as was observed for the first time in 1974. Since any electrically charged particle also has weak charge, the existence of a Z boson implies that any two electrically charged particles can interact both by exchange of photons and by exchange of Z bosons. In other words, SU(2) breaking requires electroweak unification. Page 227

Overcrossing takes place in several weak interaction processes, as shown in Figure 59. Overcrossing thus can change particle topology, and thus particle type. The strand model thus predicts that the weak interaction *changes* particle flavours (types), as is observed. Page 232

On the other hand, strands are never cut or reglued in the strand model, not even

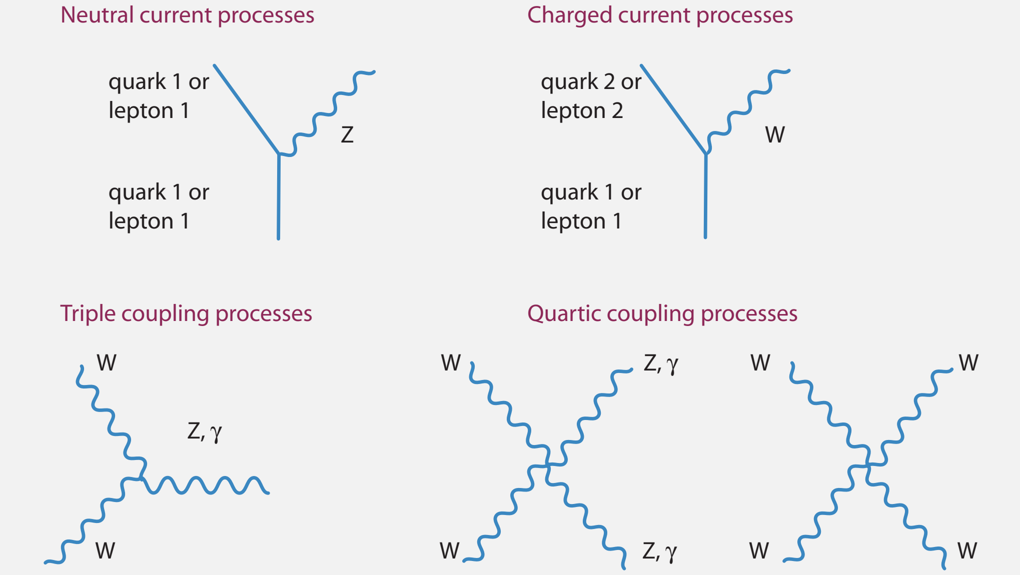


FIGURE 58 The fundamental Feynman diagrams of the weak interaction that do not involve the Higgs boson.

in the weak interaction. As a result the strand model predicts that the weak interaction conserves electric charge, spin and, as we will see below, colour charge, baryon number and lepton number. All this is observed. Page 278

Overcrossing also implies that the figure-eight knot for the Z boson and the overhand knot for the W are only the *simplest* tangles associated with each boson; more complex knots are higher order propagating states of the same basic open knots. This will be of great importance later on, for the proof that all gauge bosons of nature are already known. Page 306

In short, the second Reidemeister move leads to overcrossing; overcrossing leads to all observed properties of SU(2) symmetry breaking. The value of the mixing angle and the particle masses have still to be determined. This will be done below. Page 306

The Lagrangian of the electroweak interaction

We can now us the results on SU(2) symmetry breaking to deduce the Lagrangian density of the *electroweak* interaction. We have seen that symmetry breaking leaves the photon massless but introduces masses to the weak vector bosons, as shown in Figure 57. The result of the boson masses M_W and M_Z is that kinetic terms for the corresponding fields appear in the Lagrangian.

Due to symmetry breaking induced by overcrossing, the Z boson results from the mixing with the (unbroken) photon. The strand model predicts that the mixing can be described by an angle, the so-called weak mixing angle θ_w. In particular, the strand model implies that $\cos\theta_w = M_W/M_Z$.

As soon as symmetry breaking is described by a mixing angle due to overcrossing, we get the known Lagrangian of the electroweak interaction, though without the terms due to the Higgs boson. (We will come back to the Higgs boson later on.) We do not write down the Lagrangian of the weak interaction predicted by the strand model, but Page 286

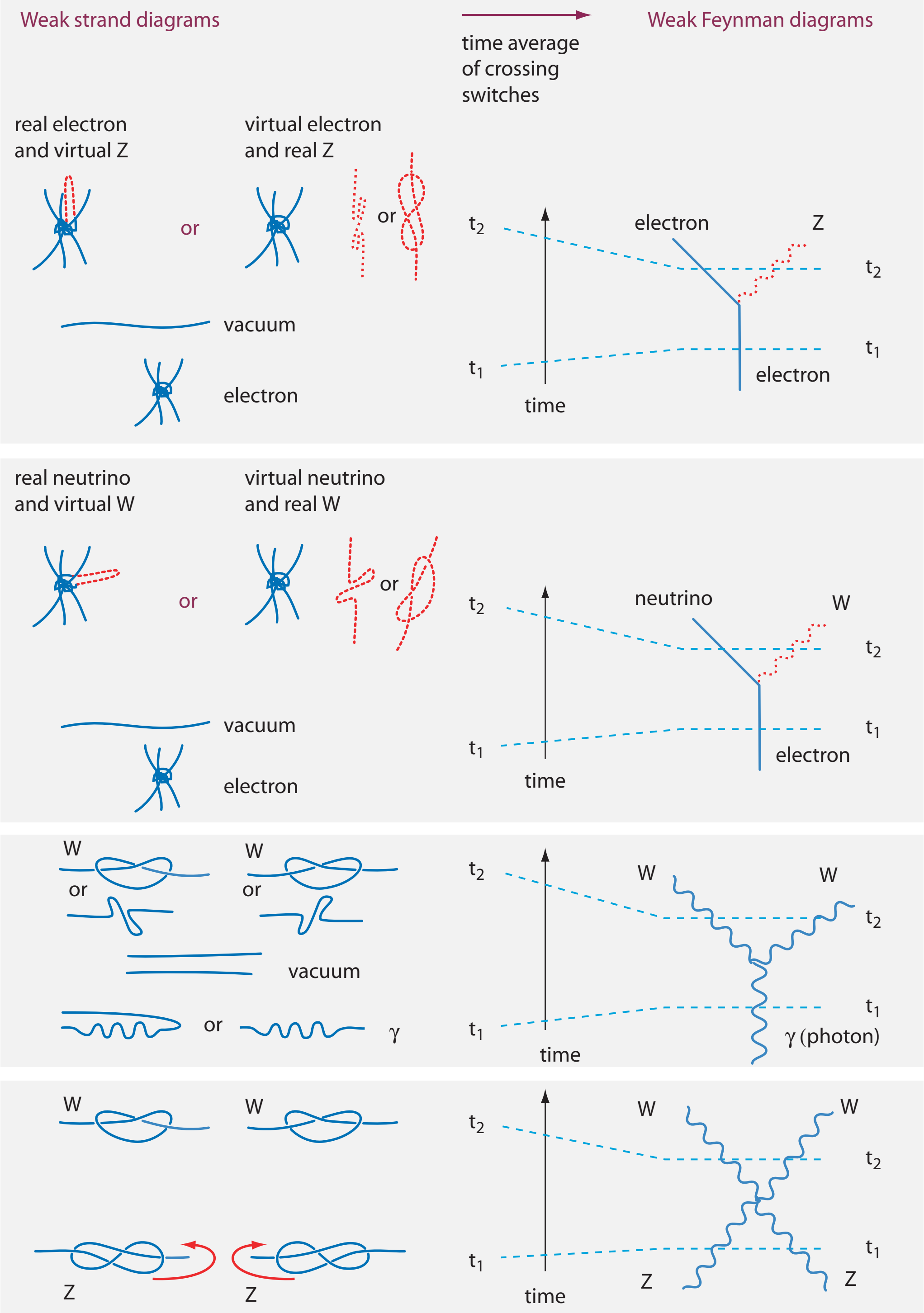


FIGURE 59 The strand model for the fundamental Feynman diagrams of the weak interaction.

the terms are the same as those found in the standard model of elementary particles. There is one important difference: the Lagrangian so derived does not yet contain the quark and lepton mixing parameters. It is found experimentally that the weak fermion eigenstates are not the same as the strong or the electromagnetic eigenstates: quarks mix, and neutrinos mix. The reason for this observation, and the effect that mixing has on the weak Lagrangian, will become clear only once we have determined the tangles for each fermion.

Page 313

In summary, the strand model implies the largest part of the Lagrangian of the weak interaction. The issue of the Higgs boson is still open, and the Lagrangian contains a number of parameters that are not yet clarified. These are the number of the involved particles, their masses mixing angles and couplings, and the value of the weak mixing angle.

The weak Feynman diagrams

In nature, the weak interaction is described by a small number of fundamental Feynman diagrams. Those not containing the Higgs boson are shown in Figure 58. These Feynman diagrams encode the corresponding Lagrangian of the weak interaction.

In the strand model, pokes lead naturally to strand versions of the fundamental Feynman diagrams. This happens as shown in Figure 59. We see again that the strand model reproduces the weak interaction: each Feynman diagram is due to a strand diagram for which only crossing switches are considered, and for which Planck size is approximated as zero size. In particular, the strand model does not allow any *other* fundamental diagrams. The small number of possible strand diagrams and thus Feynman diagrams implies that the weak interaction is *renormalizable*. For example, the running of the weak coupling with energy is reproduced by the strand model, because the running can be determined through the appropriate Feynman diagrams.

Fun challenges about the weak interaction

The W boson and its antiparticle are observed to annihilate through the electromagnetic interaction, yielding two or more photons. How can this be, given that W bosons are modelled as overhand knots?

The strand model is equivalent to gauge field theory. The strand model describes every particle as a *collection* of various tangles or knots, i.e., as a *family* of tangles or knots. This is a consequence of the properties of the weak interaction, which is able to change tangle topology. In particular, the W boson is not only a overhand knot; it also has other configurations. The most important of these other configurations, the tangle before SU(2) breaking shown in Figure 56, shows that annihilation due to electromagnetism is possible.

Page 228

Summary on the weak interaction and experimental predictions

We have deduced the main properties of the weak Lagrangian from the strand model. We have shown that pokes in tangle cores lead to a broken SU(2) gauge invariance and to massive weak bosons. We found that the deviation from tangle core sphericity plus chirality is weak charge, and that the weak interaction is non-Abelian. We have also shown that the weak interaction naturally breaks parity maximally. In short, we have deduced the main experimental properties of the weak interaction.

Is there a difference between the strand model and the description of the electroweak interaction in the standard model of particle physics? Before we can fully answer the question on deviations between the strand model and the standard model, we must settle the issue of the Higgs boson. This is done later on.

Page 286

In any case, the strand model predicts that the broken SU(2) gauge symmetry remains valid at all energies. No other gauge groups appear in nature. The strand model thus predicts again that there is no grand unification, and thus no larger gauge group, be it SU(5), SO(10), E6, E7, E8, SO(32) or any other group. This result indirectly also rules out supersymmetry and supergravity.

The strand model also predicts that the combination of gravity and quantum theory turns all Planck units into *limit* values, because there is a maximum density of strand crossings in nature, due to the fundamental principle. Therefore, the strand model predicts a *maximum weak field* value given by the Planck force divided by the smallest weak charge. All physical systems – including all astrophysical objects, such as neutron stars, quark starts, gamma ray bursters or quasars – are predicted to conform to this limit. So far, no observed field value is near this limit, so that the prediction does not contradict observation.

Our exploration of the weak interaction has left open a few points: The main points are the calculation of the weak coupling constant and the determination of the tangle for each particle of the standard model, including the Higgs. But we also need to explain the weak fermion mixings, CP violation and the masses of all particles. Despite these open points, we have settled another line of the millennium list. But before we clarify the open points, we explore the third Reidemeister move.

Page 158

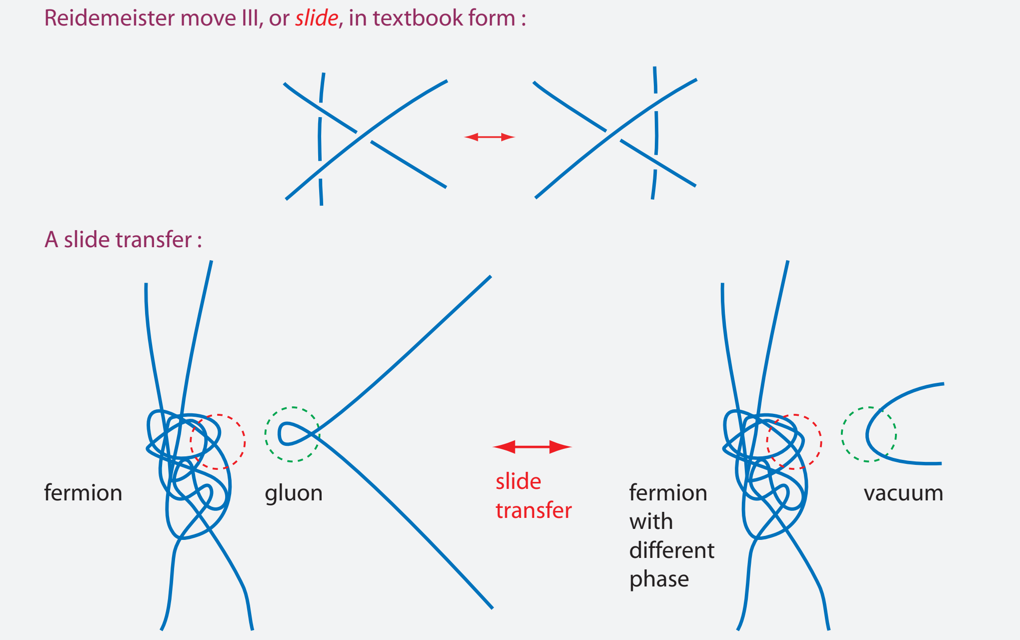


FIGURE 60 A single gluon strand changes the rotation of a tangle: *slide transfer* is the basis of the strong interaction in the strand model. No strand is cut or reglued; the transfer occurs only through the excluded volume due to the impenetrability of strands.

THE STRONG NUCLEAR INTERACTION AND THE THIRD REIDEMEISTER MOVE

In nature, the strong interaction is the result of the absorption and the emission of massless, electrically uncharged, spin-1 gauge bosons that are called *gluons*. Gluons interact with quarks, the only fermions with colour charge. Fermions can have three different colour charges, antifermions three different anticolours. Gluons form an octet, are themselves colour charged and thus also interact among themselves. The Lagrangian of quarks coupled to the gluon field has an unbroken SU(3) gauge symmetry. There are three fundamental Feynman diagrams: one for interacting quarks, and a triple and a quadruple gluon diagram. The strong coupling constant is about 0.5 at low energy; its energy dependence is fixed by renormalization.

The previous paragraph summarizes the main observations about the strong interaction. Equivalently, all observations related to the strong interaction are contained in its Lagrangian. Therefore, we need to show that the strong interaction Lagrangian follows from the strand model.

Strands and the slide, the third Reidemeister move

Page 207 As explained above, interactions of fermions are deformations of the tangle core that change its phase and rotation. We start directly by presenting the strand model for the strong interaction.

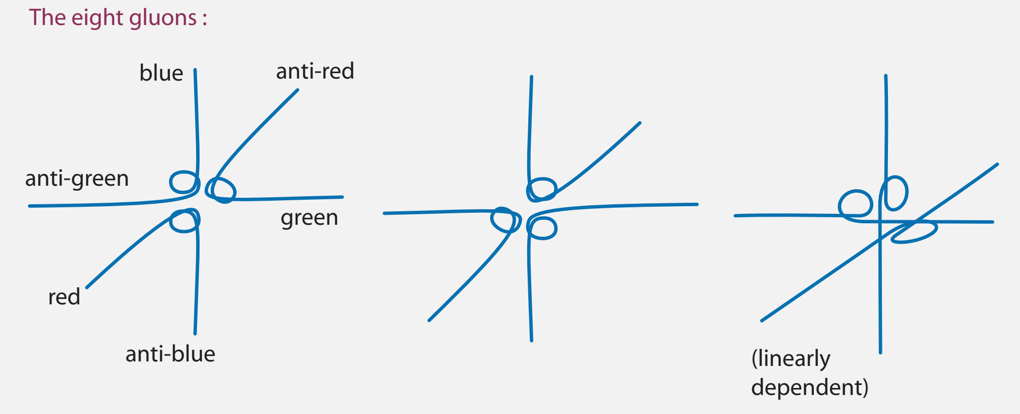


FIGURE 61 The strand model for the ‘nine’ gluons, the last three not being linearly independent.

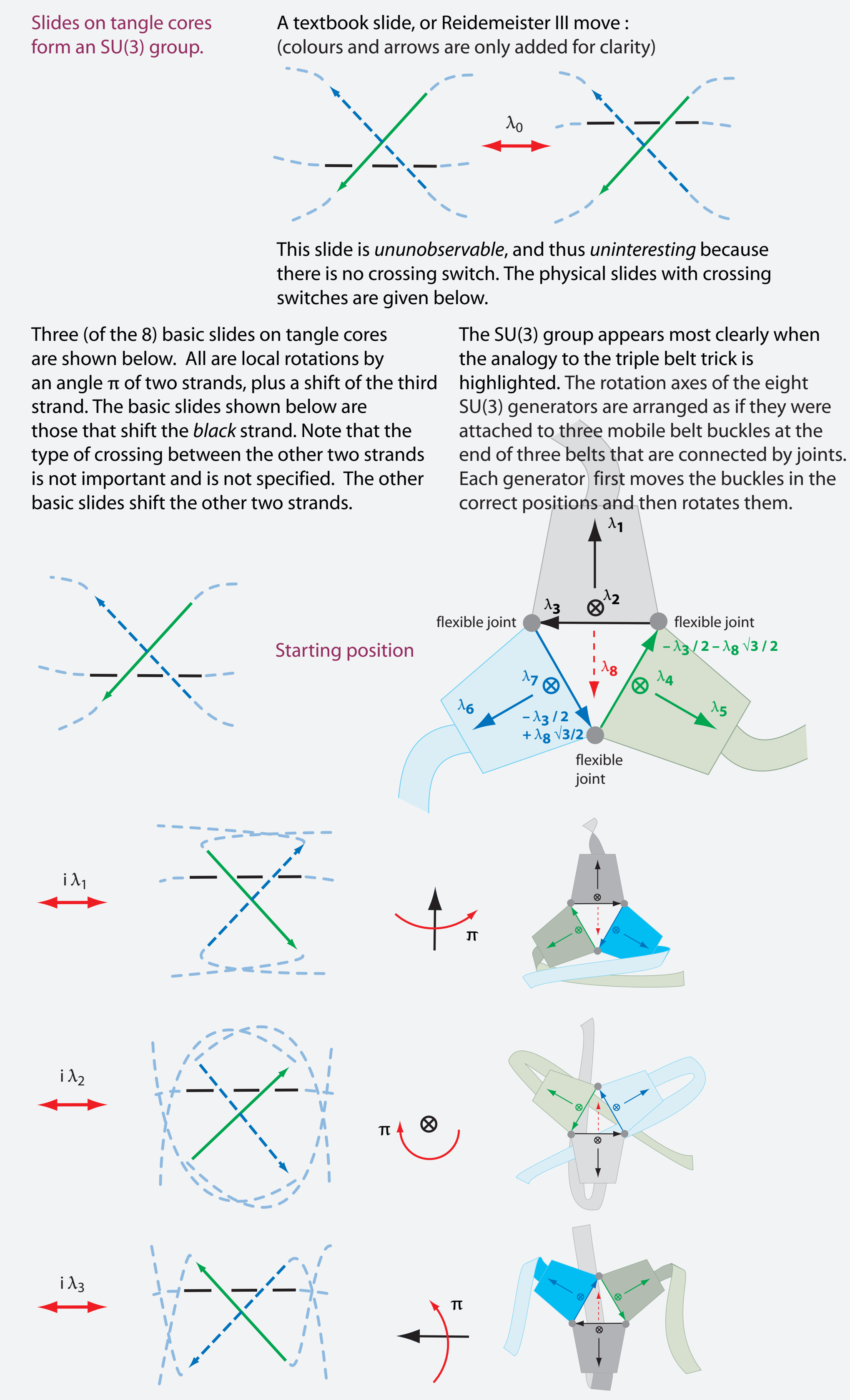
- ▷ The ***strong interaction*** is the transfer of *slides*, i.e., the transfer of third Reidemeister moves, between two particles. As shown in Figure 60, strands are not cut in this process; they simply transfer deformations as a result of their impenetrability.
- ▷ ***Gluons*** are single, unknotted strands that induce slides on colour-charged fermion cores. The strand model for gluons is shown in Figure 61.

As a result of this definition, gluons are massless, electrically neutral, and have spin 1. Due to the impenetrability of strands, a gluon strand can induce a slide on a tangle core, as shown in Figure 60. The gluon will then disappear, i.e., turn into a vacuum strand, thus leading to an effective slide transfer. This slide transfer will influence the phase of the tangle, and thus affect the rotation of the tangle. This shows that slide transfers are indeed a type of interaction.

From slides to SU(3)

To find out what algebraic structure is generated by slides, we must explore them in detail. A slide, or third Reidemeister move, involves three pieces of strands. The textbook version of the slide, called λ_0 in Figure 62, deforms strands by sliding one strand, drawn in black in the figure, against a crossing of the other two. However, such a deformation does not contain any crossing switch; following the fundamental principle of the strand model, it is therefore unobservable, or, simply said, of no physical relevance. However, similar deformations that do involve crossing switches also exist. They are called λ_1 to λ_3 in Figure 62. These deformations involve combined rotations by π of two strands, thus involve crossing changes, and therefore are physical.

We note that ‘slide’ is not a good term for these operations; in fact, they are combinations of a local *rotation* by π, a *flattening* to the observation plane, and a *slide*. Nevertheless, we will continue to call them ‘slides’ for brevity. We also note that the generalized, observable slides just defined *differ* from twists and pokes, because they require three pieces of strands and because they are constructed with the help of the (unphysical) λ_0 move.



Slides on tangle cores form an SU(3) group.

A textbook slide, or Reidemeister III move : (colours and arrows are only added for clarity)

This slide is *ununobservable*, and thus *uninteresting* because there is no crossing switch. The physical slides with crossing switches are given below.

Three (of the 8) basic slides on tangle cores are shown below. All are local rotations by an angle π of two strands, plus a shift of the third strand. The basic slides shown below are those that shift the *black* strand. Note that the type of crossing between the other two strands is not important and is not specified. The other basic slides shift the other two strands.

The SU(3) group appears most clearly when the analogy to the triple belt trick is highlighted. The rotation axes of the eight SU(3) generators are arranged as if they were attached to three mobile belt buckles at the end of three belts that are connected by joints. Each generator first moves the buckles in the correct positions and then rotates them.

FIGURE 62 The strand deformations for the eight gluons λ_1 to λ_8 associated to the slide move, and the SU(3) structure of the strong interaction as the result of three joined belts.

The observable slide moves λ_1, λ_2 and λ_3 can be concatenated. The concatenation of any of these three moves with itself yields a situation that can (almost) be described as the original region after rotation by 2π. We saw above that such a deformation corresponds to a multiplication of the original situation by -1. When two of these slide moves are concatenated, we get the same result as the third (up to a sign that depends on whether the combination is cyclical or not). Thus the three moves form an SU(2) group, one of the several subgroups of SU(3). Therefore they can be visualized by the three orthogonal rotations by π of the buckle at the end of a belt. Page 177

The observable slide moves may involve slides of either of three strands. This yields a total of 9 observable slides for the observer defined by the paper plane. The slides corresponding to λ_1 are usually called λ_5 and λ_6, those to λ_2 are called λ_4 and λ_7. Three of the slides constructed in this way are linearly dependent, namely those formed by λ_3 and its two 'cousins'; among them, only *two* are needed. They are called λ_3 and λ_8 in the Gell-Mann set and in Figure 62. In total, this gives 8 linearly independent slides.

As just mentioned, the slides λ_1, λ_2 and λ_3 form an SU(2) subgroup. The same is true for the corresponding slides: also the triplet λ_5, λ_4, $-\lambda_3/2 - \lambda_8\sqrt{3}/2$ and the triplet λ_6, λ_7, $-\lambda_3/2 + \lambda_8\sqrt{3}/2$ form SU(2) groups. These three SU(2) groups are linearly independent and each of them can be represented by a belt. We thus deduce that the eight slides can be represented by three belts whose buckles are connected by flexible joints, as shown in Figure 62.

The visualization with three belts allows us to explore slide concatenations by playing with a physical structure. A few details are important. First, the slides of Figure 62 correspond to i times the Gell-Mann generators. Second, all slides are combinations of rotations and flattenings. The starting position of the system has the three buckles at right angles to each other. For this reason, the square of a slide is not really -1, but involves λ_8 and/or λ_3. Third, multiplying slides is concatenation, whereas adding slides is the operation defined above: addition is the result of connecting partial tangles. Page 177

We can now play with the triple belt structure and read off the multiplication behaviour of the eight generators. The multiplication table is:

$\cdot$	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
λ_1	$2/3 + \lambda_8/\sqrt{3}$	$i\lambda_3$	$-i\lambda_2$	$i\lambda_7/2 + \lambda_6/2$	$-i\lambda_6/2 + \lambda_7/2$	$i\lambda_5/2 + \lambda_4/2$	$-i\lambda_4/2 + \lambda_5/2$	$+\lambda_1/\sqrt{3}$
λ_2	$-i\lambda_3$	$2/3 + \lambda_8/\sqrt{3}$	$i\lambda_1$	$i\lambda_6/2 - \lambda_7/2$	$i\lambda_7/2 + \lambda_6/2$	$-i\lambda_4/2 + \lambda_5/2$	$-i\lambda_5/2 - \lambda_4/2$	$+\lambda_2/\sqrt{3}$
λ_3	$i\lambda_2$	$-i\lambda_1$	$2/3 + \lambda_8/\sqrt{3}$	$i\lambda_5/2 + \lambda_4/2$	$-i\lambda_4/2 + \lambda_5/2$	$-i\lambda_7/2 - \lambda_6/2$	$i\lambda_6/2 - \lambda_7/2$	$+\lambda_3/\sqrt{3}$
λ_4	$-i\lambda_7/2 + \lambda_6/2$	$-i\lambda_6/2 - \lambda_7/2$	$-i\lambda_5/2 + \lambda_4/2$	$2/3 + \lambda_3/2 - \lambda_8/2\sqrt{3}$	$i\lambda_3/2 + i\sqrt{3}\lambda_8/2$	$i\lambda_2/2 + \lambda_1/2$	$i\lambda_1/2 - \lambda_2/2$	$-i\sqrt{3}\lambda_5/2 - \lambda_4/2\sqrt{3}$
λ_5	$i\lambda_6/2 + \lambda_7/2$	$-i\lambda_7/2 + \lambda_6/2$	$i\lambda_4/2 + \lambda_5/2$	$-i\lambda_3/2 - i\sqrt{3}\lambda_8/2$	$2/3 + \lambda_3/2 - \lambda_8/2\sqrt{3}$	$-i\lambda_1/2 + \lambda_2/2$	$i\lambda_2/2 + \lambda_1/2$	$i\sqrt{3}\lambda_4/2 - \lambda_5/2\sqrt{3}$
λ_6	$-i\lambda_5/2 + \lambda_4/2$	$i\lambda_4/2 + \lambda_5/2$	$i\lambda_7/2 - \lambda_6/2$	$-i\lambda_2/2 + \lambda_1/2$	$i\lambda_1/2 + \lambda_2/2$	$2/3 - \lambda_3/2 - \lambda_8/2\sqrt{3}$	$-i\lambda_3/2 + i\sqrt{3}\lambda_8/2$	$-i\sqrt{3}\lambda_7/2 - \lambda_6/2\sqrt{3}$
λ_7	$i\lambda_4/2 + \lambda_5/2$	$i\lambda_5/2 - \lambda_4/2$	$-i\lambda_6/2 - \lambda_7/2$	$-i\lambda_1/2 - \lambda_2/2$	$-i\lambda_2/2 + \lambda_1/2$	$i\lambda_3/2 - i\sqrt{3}\lambda_8/2$	$2/3 - 2\lambda_3 - \lambda_8/2\sqrt{3}$	$i\sqrt{3}\lambda_6/2 - \lambda_7/2\sqrt{3}$
λ_8	$+\lambda_1/\sqrt{3}$	$+\lambda_2/\sqrt{3}$	$+\lambda_3/\sqrt{3}$	$i\sqrt{3}\lambda_5/2 - \lambda_4/2\sqrt{3}$	$-i\sqrt{3}\lambda_4/2 - \lambda_5/2\sqrt{3}$	$i\sqrt{3}\lambda_7/2 - \lambda_6/2\sqrt{3}$	$-i\sqrt{3}\lambda_6/2 - \lambda_7/2\sqrt{3}$	$2/3 - \lambda_8/\sqrt{3}$

As expected, this table is the algebra of the Gell-Mann matrices, which form a standard set of generators of the group SU(3). We have thus deduced that the eight linearly independent slides that can be applied to a tangle core represent the eight virtual gluons that can act on a fermion. The table proves that the 8 gluons transform according to the adjoint (and faithful) representation of SU(3), as is stated by quantum chromodynamics. In other words, we have shown that the Lagrangian of strongly interacting fermions has a SU(3) gauge invariance.

The slide analogy for gluons implies that gluons are described by single unknotted strands that impart 'slides' to fermions. A simple image is to describe real gluons as *loops* that 'pull' one strand during the slide, as shown in Figure 61. This single strand model also reproduces the vanishing mass of gluons and their spin 1 value.

The 8 gluons transform according to the adjoint (and faithful) representation of SU(3). The multiplication table shows that two general slides do not commute and do not anticommute. The group SU(3) is *non-Abelian*. This implies that gluons interact among themselves. Both the multiplication table and the strand model for gluons implies that two interacting gluons can yield either one or two gluons, but not more, as illustrated in Figure 63. Since in the strand model, gluons do not change topology, but only shapes, gluons are predicted to be massless, despite interacting among themselves. Since gluons interact among themselves, free gluons do not appear in nature.

Slides, i.e., gluon emission or absorption, never change the topology of tangles. Thus the strand model predicts that the strong interactions conserve electric charge, baryon number, weak isospin, flavour, spin and all parities. This is indeed observed. In particular, there is a natural lack of CP violation by slides. This is precisely what is observed

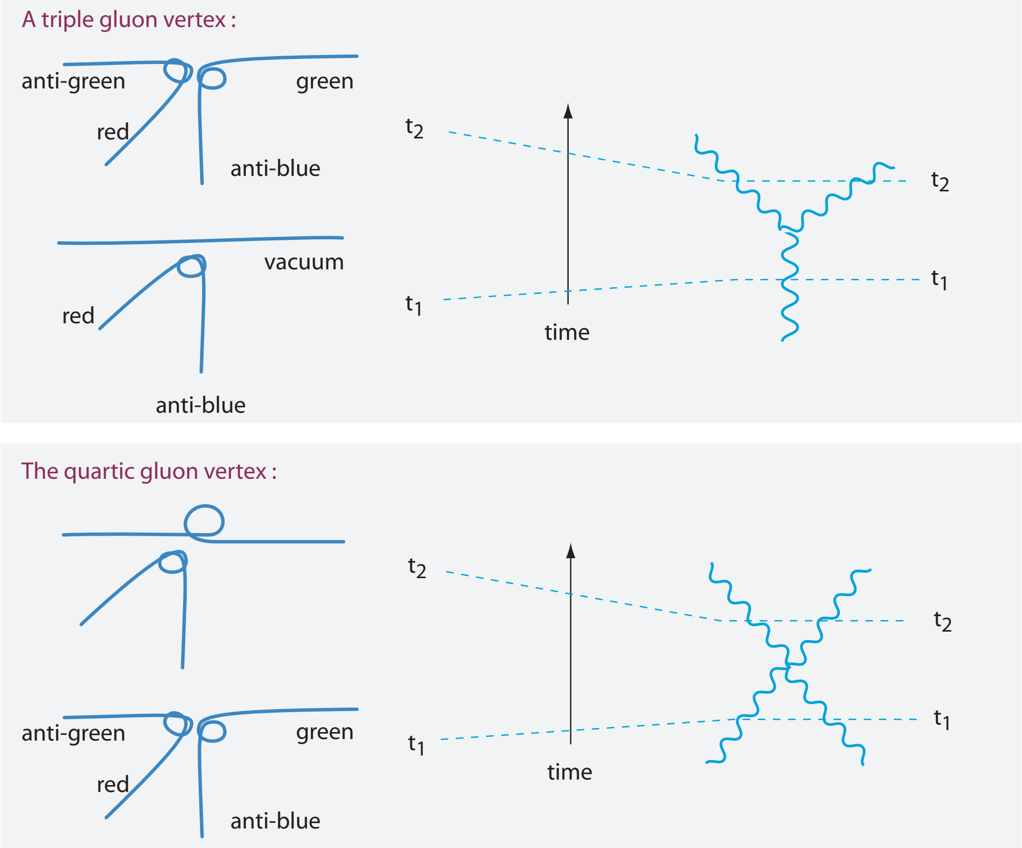


FIGURE 63 The self-interaction of gluons in the strand model.

about the strong interaction. The strand model thus also reproduces the lack of CP violation by the strong interaction.

Open challenge: Find a better argument for the gluon tangle

Challenge 133 ny

The argument that leads to the gluon tangle is too hand-waving. Can you give a better argument?

The gluon Lagrangian

Gluons are massless particles with spin 1. As a result, the field intensities and the Lagrangian are determined in the same way as for photons: energy density is the square of crossing density, i.e., the 'square' of field intensity. Since there are 8 gluons, the Lagrangian density becomes

$$\mathcal{L}_{\text{gluons}} = -\frac{1}{4} \sum_{a=1}^{8} G^{a}_{\mu\nu} G_{a}^{\mu\nu} \qquad (170)$$

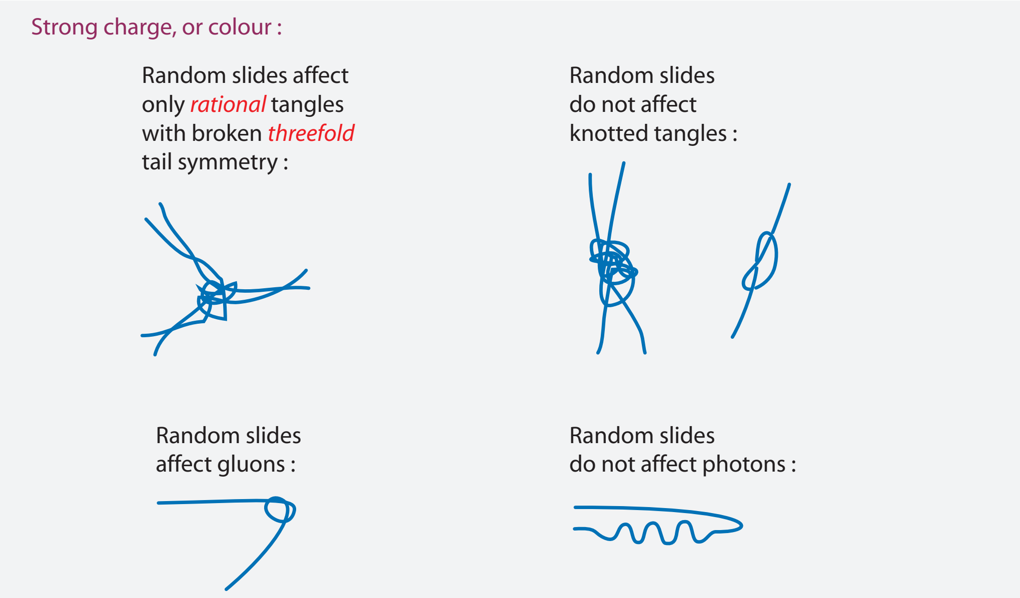


FIGURE 64 Tangles with and without colour charge.

where the gluon field intensities, with *two* greek indices, are given naturally as

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g f^{abc} G^b_\mu G^c_\nu \; , \tag{171}$$

and f_{abc} are the structure constants of SU(3) that can be deduced from the multiplication table given above. The quantities G^a_μ, with *one* greek index, are the gluon vector potentials. The last term in the definition of the field intensities corresponds to the diagram of Figure 63. The Lagrangian is simply the natural generalization from the U(1) case of Page 215 photons to the SU(3) case of gluons. In short, we obtain the usual free gluon Lagrangian from the strand model.

Colour charge

Surrounded by a bath of gluons that randomly induce slides of all kinds, not all fermion cores will change their rotation state. Generally speaking, particles have colour if a bath of random gluons changes their phase. Only tangles which lack some symmetry will have colour charge. Tangle which are symmetric will be neutral, or 'white'. Which symmetry is important here?

We see directly that the photon tangle is not sensitive to a gluon bath. The same is valid for W and Z bosons. The strand model predicts that they are colour-neutral, i.e., that they are 'white', as is observed.

As just mentioned, gluons interact among themselves. It is usual to say that gluons have a colour and an anticolour, as this is one way to describe the representation to which they belong.

Let us now explore fermions. In the strand model, a fermion has colour charge if the

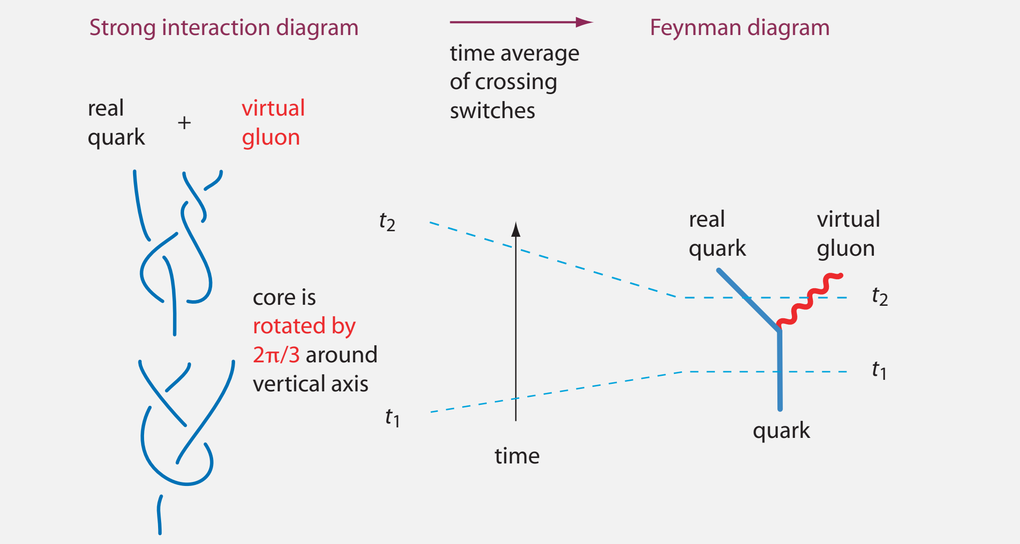


FIGURE 65 The Feynman diagram of the strong interaction for a quark. The upper triplet of tails correspond to the three belts.

corresponding triple belt model is affected by large numbers of random gluons. The first tangles that come to mind are tangles made of three strands, such as the simple tangles shown in Figure 62. But a short investigation shows that such tangles are colour-neutral, or 'white'. In contrast, a rational tangle made of *two* strands does not suffer this fate. For a bath of boson strands that induce slides, i.e., third Reidemeister moves, general *rational tangles* made of two strands are expected to be influenced, and thus to be colour-charged.

Rational tangles made of two strands are the simplest possible tangles. An example is shown in Figure 65. Such tangles break the three-fold symmetry of the three-belt structure, and are thus colour-charged. We will show below how these tangles are related to quarks.

▷ A fermion tangle has ***colour charge*** if its three-belt model is not symmetric for rotations by ±2π/3.

Coloured rational tangles automatically have *three* possible colours:

▷ The ***three colour charges*** are the three possibilities to map a tangle to the three belt model.*

Each colour is thus a particular orientation in ordinary space.

If we explore other types of tangles made of two strands, such as *prime* tangles or *intrinsically knotted* tangles, we find that their colour depends on their structure. If the knots are reduced with overcrossing as much as possible, the colour of such a complex tangle is the colour of the rational tangle that we obtain after the reduction. The strand

---

Challenge 134 ny

* Can you define a knot invariant that reproduces colour charge? And a geometric knot invariant that does so?

model thus predicts that rational tangles made of two strands are the basic colour states. And indeed, in nature, quarks are the only fermions with colour charge.

Properties of the strong interaction

In the strand model, all interactions are deformations of the tangle core. The strong interaction is due to exchange of slides. In the case of coloured fermions, colour change is a change of the mapping to the three-belt model, i.e., a change of orientation of the tangle in space.

Visual inspection shows that slide exchanges, and thus gluon exchanges, *conserve* colour. Since the strong interaction conserves the topology of all involved tangles and knots, the strong interaction also conserves electric charge, parity, and, as we shall see below, all other quantum numbers. All these results correspond to observation.

The Lagrangian of QCD

Starting from the idea that tangle core deformations lead to phase redefinitions, we have thus found that slides imply that the complete strong interaction Lagrangian density for matter and radiation fields is SU(3) gauge invariant. If we insert this gauge invariance into the fermion Lagrangian density, we get

$$\mathcal{L}_{\mathrm{QCD}} = \sum_q \overline{\Psi}_q \left(i\hbar c \not{D} - m_q c^2 \delta_{qq'} \right) \Psi_{q'} - \frac{1}{4} \sum_{a=1}^{8} G^a{}_{\mu\nu} G_a{}^{\mu\nu} \ , \qquad (172)$$

where the index q counts the coloured fermion, i.e., the quark. In the Lagrangian density, $\not{D}$ is now the SU(3) gauge covariant derivative

$$\not{D} = \not{\partial} - g\,\gamma^{\mu} G^{a}_{\mu} \lambda_a \ , \qquad (173)$$

where g is the gauge coupling, λ_a are the generators of SU(3), i.e., the Gell-Mann matrices given above, and the G^a_μ are, as before, the gluon vector potentials. The last term in the covariant derivative corresponds to the Feynman diagram and the strand diagram of Figure 65.

In summary: the strand model reproduces QCD. However, we have not yet deduced the number and masses m_q of the quarks, nor the gauge coupling g.

Renormalization of the strong interaction

The slide move implies that only three Feynman diagrams appear in strong nuclear reactions: only one QCD Feynman diagram exists for quarks, and only triple and quadruple vertices exist among gluons. This limited range of options is essential for the renormalization of QCD and allowed us to deduce the QCD Lagrangian. The strand model thus automatically ensures that the strong interaction is renormalizable.

The strand model provides a new underlying picture for the Feynman diagrams of the strong interaction, but does not change the physical results at any energy scale accessible in the laboratory. In particular, the running of the strong coupling constant is reproduced. Indeed, in the strand model, a flux-tube–like bond between the quarks appears automatically, as we will see when exploring hadrons. At high kinetic energies, the bond has little effect, so that quarks behave more like free particles. In short, the strand model reproduces asymptotic freedom. An argument for quark confinement will be given later on.

Page 290, page 296

Page 293

CURIOSITIES AND FUN CHALLENGES ABOUT SU(3)

Vol. V, page 212

Deducing the Lie groups SU(3) – and SU(2) – from deformations is a new result. Frank Wilczek, Alfred Shapere, Alden Mead, Jerry Marsden and several others have confirmed that so far, only the geometric Lie group SO(3) or its subgroups had been found in deformations. The fundamental postulate shows its power by overcoming this limitation.

Ref. 183

Deducing the Lie groups U(1), SU(2) and SU(3) directly from a basic principle contradicts another old dream. Many scholars hoped that the three gauge groups have something to do with the sequence complex numbers, quaternions and octonions. The strand model quashes this hope.

* *

The Lie group SU(3) is also the symmetry group of the three-dimensional harmonic oscillator. What is the relation to the Lie group SU(3) induced by slides?

Challenge 135 ny

* *

Show that the strand model does not contradict the Coleman–Mandula theorem on the possible conserved quantities in quantum field theory.

Challenge 136 ny

* *

Show that the strand model does not contradict the Weinberg–Witten theorem on the possible massless particles in quantum field theory.

Challenge 137 ny

* *

Are the *Wightman axioms* of quantum field theory fulfilled by the strand model *with* interactions? The *Haag–Kastler axioms*?

Challenge 138 ny

* *

Ref. 184

Show that the BCFW recursion relation for tree level gluon scattering follows from the strand model.

Challenge 139 ny

SUMMARY ON THE STRONG INTERACTION AND EXPERIMENTAL PREDICTIONS

Is there a difference between the strand model and QCD? Not as long as gravity plays no role. The strand model predicts that this will only happen near the Planck energy $\sqrt{\hbar c^5/4G}$. In particular, no other gauge groups appear. The strand model thus predicts again that there is no grand unification in nature, and thus no larger gauge group. Often discussed groups such as SU(5), SO(10), E6, E7, E8 or SO(32) are predicted not to apply to nature. So far, this prediction agrees with experiment.

The strand model also predicts that the strong interaction is naturally CP-invariant. This means that axions, particles invented to explain the invariance, are unnecessary: as shown below, the strand model even predicts that they do not to exist. So far, both predictions agree with experiment. Page 304

The strand model predicts that the combination of gravity and quantum theory turns all Planck units into *limit* values. The strand model thus predicts a maximum strong field value given by the Planck force divided by the strong charge of the quark. All physical systems – including all astrophysical objects, such as neutron stars, quark stars, gamma ray bursters or quasars – are predicted to conform to this field limit. So far, this prediction is validated by experiment.

In summary, we have deduced the Lagrangian density of QCD from the strand model. We have shown that slides in tangle cores lead to an SU(3) gauge invariance, and that strong charge is related to the topology of certain rational tangles. In short, we have deduced most observed properties of the strong interaction. We have not yet deduced the tangles and the number of quarks, their masses and the strength of the coupling. Despite these open issues, we have settled another issue of the millennium list. Page 158

SUMMARY ON MILLENNIUM ISSUES: GAUGE INTERACTIONS

At this point, we have deduced quantum field theory and the three gauge interactions from strands. Doing this, we explained the dimensions of space-time, the Planck units, the principle of least action, the appearance of gauge groups, of renormalization, of Lorentz symmetry and of permutation symmetry. Thus we have deduced all the concepts and all the mathematical structures that are necessary to *formulate* the standard model of elementary particles.

We have not yet deduced the full standard model: we still need to show which particles exist, which properties they have and what couplings they produce. However, we have found that the strand model explains all the mathematical structures from the millennium list that occur in quantum field theory. This explanation also allows to state what interactions do *not* occur in nature. Page 158

PREDICTION ABOUT THE NUMBER OF INTERACTIONS

Already in 1926, Kurt Reidemeister proved an important theorem about possible deformations of knots or tangles that involve crossing switches. When tangles are described with two-dimensional diagrams, all possible deformations can be reduced to exactly three moves, nowadays called after him. In the strand model, the two-dimensional tangle diagram describes what an observer *sees* about a physical system. Reidemeister's theorem, together with the equivalence of interactions as crossing-changing deformations, thus proves that there are *only three gauge interactions* in nature. Searches for other gauge interactions are predicted to fail, as they have up to now. Ref. 182

UNIFICATION OF INTERACTIONS

Ref. 145 On the other hand, we can also state that there is only *one* Reidemeister move. This becomes especially clear if we explore the three-dimensional shape of knots, instead of their two-dimensional diagrams: all three Reidemeister moves can be deduced from the *same* deformation a single strand. Only the projection on a diagram creates the distinction between the three moves. In the terms of the strand model, this means that all gauge interactions are in fact aspects of only one basic process, a fluctuation of strand shape, and that the three gauge interactions are distinguished by the diagram view. In this way, the three gauge interactions are thus *unified* by the strand model.

In the strand model, the projection plane maps strand fluctuations to Reidemeister moves. The projection plane is defined by the observer, i.e., by the frame of reference. In short, the type of interaction depends on the observer. In nature, however, this is not the case. This apparent contradiction can be solved. In the strand model, the type of interaction of a particle results from the type of asymmetry of its tangle. Other strand deformations do not lead to interactions, because their effects are strongly suppressed by the averaging of short-time fluctuations underlying every observation. In short, the averaging process at the basis of interactions also ensures that interactions are effectively observer-independent.

Predictions about grand unification and supersymmetry

The three gauge interactions are due to the three Reidemeister moves. Therefore, the strand model asserts that there is no single gauge group for all interactions: there is no so-called *grand unification*. The absence of grand unification implies the absence of large proton decay rates, the absence of additional, still undiscovered gauge bosons, the absence of neutron–antineutron oscillations, and the absence of sizeable electric dipole moments in elementary particles. All these searches are ongoing at present, and are predicted to yield no results.

Reidemeister moves are confined to three spatial dimensions. Indeed, the strand model is based on exactly three spatial dimensions. There are no other, undetected dimensions; three-dimensional space is the time average of unlinked strands. There is no supersymmetry and no supergravity. The strand model thus predicts the absence of all conjectured 'superparticles'. The strand model also predicts the absence of non-commutative space-time, even though, with some imagination, strands can be seen as remotely related to that approach. In short, the strand model differs experimentally and theoretically from the unification proposals made in the twentieth century.

No new observable gravity effects in particle physics

We can summarize our findings also in the following way: *the strand model predicts that masses are the only observable effect of gravity in particle physics.* This result will be complemented below by a second, equally restrictive result that limits the observable quantum effects in the study of gravity. In short, the strand model keeps particle physics and general relativity almost completely separated from each other.

Page 272

The status of our quest

In this chapter, we have deduced that strands predict exactly three interactions, because interactions are deformations of tangle cores and because there are only three classes of core deformations, given by the three Reidemeister moves. The three interactions are described by a U(1), a broken SU(2) and a SU(3) gauge symmetry, respectively. Strands also show that the three interactions are renormalizable, relativistically invariant, and that they follow the least action principle. Strands thus imply the interaction Lagrangians of the standard model of particle physics.

Page 158

If we look at the millennium list of open issues in fundamental physics, we have now solved all issues about the mathematical structures that appear in quantum field theory and in the standard model of particle physics. Two groups of issues are still unexplained: the number and properties of the elementary particles and the theory of general relativity. We start with the latter.

Chapter 11

GENERAL RELATIVITY DEDUCED FROM STRANDS

General relativity describes the deformations of the vacuum. In everyday life, gravitation is the only such effect that we observe. But on astronomical scale, gravity shows more phenomena: vacuum can deflect light, producing gravitational lenses, can wobble, giving gravitational waves, and can accelerate, yielding the darkness of the sky and the fascinating black holes. All these observations require general relativity for their description. Therefore, general relativity must be part of any unified description of nature. The next task is thus set: to deduce the field equations of general relativity from the strand model.

Flat space, special relativity and its limitations

We have seen above that any observer automatically introduces a 3+1-dimensional *background* space-time. We have also seen that in the case of quantum theory, *physical* space-time, the space-time that is formed by the fluctuations of the vacuum strands, is naturally 3+1-dimensional and flat. In the absence of gravity, physical space and background space coincide.

Page 192

Using strands, we have deduced the invariant limit c for all energy speeds and shown that it is realized only for free massless particles, such as photons. Strands also showed us that massive particles move more slowly than light. In short, strands reproduce special relativity.

Page 195

The strand model thus predicts that *pure* special relativity is correct for all situations and all energies in which gravity and quantum theory plays no role. The strand model also predicts that when gravity or quantum effects do play a role, general relativity or quantum theory *must* be taken into account. This means that there is no domain of nature in which intermediate descriptions are valid.

It is sometimes suggested that the invariant Planck energy limit for elementary particles might lead to a 'doubly special relativity' that deviates from special relativity at high particle energy. However, this suggestion is based on two assumptions: that at Planck energy *point masses* are a viable approximation to particles, and that at Planck energy *vacuum and matter differ*. In nature, and in the strand model, both assumptions are incorrect. Nature, as general relativity shows, does not allow the existence of point masses: the densest objects in nature are black holes, and these are not point-like for any mass value. Furthermore, at Planck energy, matter and vacuum *cannot* be distinguished. Put simply, no system at Planck energy can be described without general relativity and quantum gravity. In short, the strand model predicts that the approach of 'doubly special rel-

Ref. 84
Page 59
Ref. 185

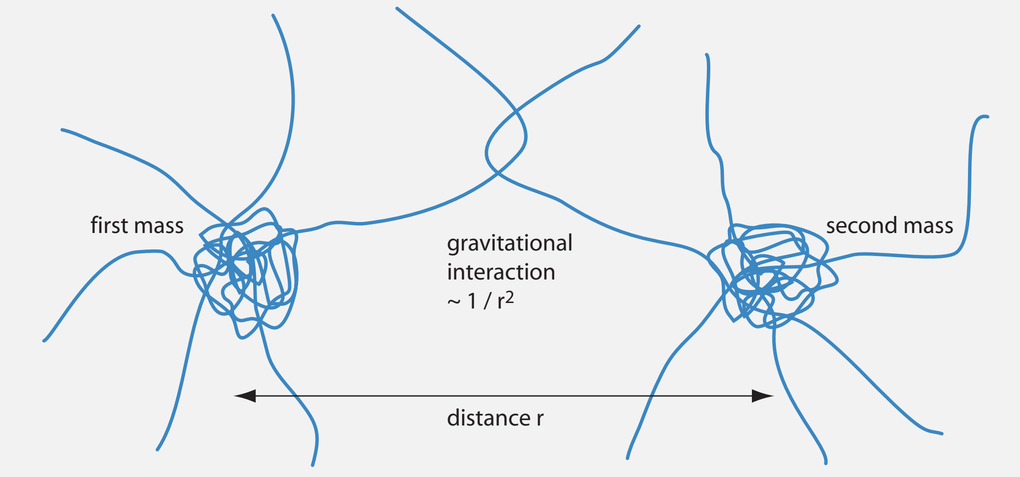


FIGURE 66 Gravitational attraction as result of strands.

ativity' cannot be correct.

Classical gravitation

In nature, at low speeds and in the flat space limit, gravitation is observed to lead to an acceleration of test masses that changes as the inverse square distance from the gravitating mass. This acceleration is called *universal gravitation* or *classical gravitation*. It is an excellent approximation for the solar system and for many star systems throughout the universe.

In the most common view, *gravitation appears because any mass M generates vacuum energy around it.* The strand model reproduces this connection. In the strand model, every space-time effect, including gravitation, is due to the behaviour of tangle tails. In the strand model, every mass, i.e. every system of tangles, is connected to the border of space by tails. The nearer a mass is to a second mass, the more frequently the tails of the two masses cross and get tangled. Figure 66 illustrates the situation. The strand model states:

▷ ***Gravitation*** is due to the fluctuations of tail crossings.

The tail crossings fluctuate; averaged of time, the fluctuations lead to a crossing switch density. The resulting energy density – where energy is the number of crossing switches per time – changes like the inverse distance from the central mass. This is the reason for the $1/r$-dependence of the gravitational potential and the $1/r^2$-dependence of gravitational acceleration. (This applies to all those cases where curvature is negligible.) In simple words, in the strand model, the inverse square dependence of gravitational acceleration is due to the three-dimensionality of space combined with the one-dimensionality of strands. Challenge 140 e

The strand model also shows that masses and energies are always positive: every tangle has tails. The model also shows qualitatively that larger masses produce stronger at-

Page 306 traction, as they generally have more tails. We will show below that the effective number density of tails is indeed proportional to the mass.

In the strand model, crossing switches are not only related to energy; they are also related to entropy. A slightly different – but equivalent – view on gravitation therefore appears when we put the stress on the entropic aspect. An especially clear explanation Ref. 186 was recently given by Erik Verlinde. In this view, *gravity appears because any mass M generates an effective vacuum temperature around it.* A gravitating mass M attracts test masses because during the *fall* of a test mass, the total entropy *decreases*. It is not hard to describe these ideas quantitatively.

Given a spherical surface A enclosing a gravitating mass M at its centre, the acceleration a of a test mass located somewhere on the surface is given by the local vacuum temperature T:

$$a = T\,\frac{2\pi\,kc}{\hbar}\ , \tag{174}$$

where k is the Boltzmann constant. This relation is called the *Fulling–Davies–Unruh effect* and relates vacuum temperature and local acceleration.* In the strand model, the vacuum temperature at the surface of the enclosing sphere is given by the crossing switches that the tails starting at the mass induce there. We can determine the vacuum temperature by dividing the energy E contained inside the sphere by *twice* the maximum possible entropy S for that sphere. This maximum value is the entropy the sphere would have if it were a black hole horizon and can be calculated by the strand model, as we will see Page 255 shortly. This yields the expression

$$T = \frac{E}{2S} = \frac{M}{A}\,\frac{2G\hbar}{kc}\ . \tag{175}$$

Neglecting spatial curvature, we can set $A = 4\pi R^2$; this gives a temperature at the enclosing sphere given by

$$T = \frac{M}{R^2}\,\frac{G\hbar}{2\pi\,kc}\ . \tag{176}$$

Inserting this expression into the expression for the Fulling–Davies–Unruh acceleration, we get

$$a = G\,\frac{M}{R^2}\ . \tag{177}$$

This is the law of universal gravitation. Since spatial curvature was neglected, and the central mass was assumed at rest, this expression is only valid for large distances and small speeds. We have thus deduced universal gravity from the effects of gravitating masses on vacuum temperature. Below, we generalize this sequence of arguments to the Page 259 relativistic case and deduce the field equations of general relativity.

Alternatively, it can be argued that the gravitational force F on a test mass m is given by the vacuum temperature created by the central mass M and by the change of entropy

---

* An inertial or a freely falling mass (or observer) thus measures a vanishing vacuum temperature.

S per length that is induced by the motion of the test mass:

$$F = T \frac{\mathrm{d}S}{\mathrm{d}x} . \tag{178}$$

The change of entropy $\mathrm{d}S/\mathrm{d}x$ when a test mass m is moved by a distance x can be determined from the strand model. When a mass m moves by a (corrected) Compton length, in the strand model, the mass has rotated by a full turn: the entropy change is $2\pi k$ per (corrected) Compton length. Thus we have

$$\frac{\mathrm{d}S}{\mathrm{d}x} = m \frac{2\pi\, kc}{\hbar} . \tag{179}$$

Using the temperature found above, we get an expression for the gravitational force given by

$$F = G \frac{Mm}{R^2} . \tag{180}$$

This is the force law of universal gravitation, as discovered by Robert Hooke and popularized by Isaac Newton. We have thus deduced universal gravity from the entropy generated by gravitating masses.

The last two arguments leading to universal gravitation started from black hole entropy. In the strand model, black hole entropy is a consequence of the underlying strand crossing switches. Universal gravitation thus (again) appears as an effect of the crossing switches induced by masses. Page 255

In summary, independently of the assumed viewpoint, strands do explain the origin of universal gravitation. Incidentally, modelling mass as a source for strand crossing switches is remotely reminiscent of Georges-Louis Lesage's eighteenth-century model of gravitation. Lesage proposed that gravity appears because many tiny, usually unnoticed corpuscules push masses together. Ref. 187 In fact, as we will see shortly, there is a similarity between these assumed tiny corpuscules and virtual gravitons. And interestingly, all criticisms of Lesage's model then cease to hold. Vol. I, page 176 First, there is no deceleration of free masses in inertial motion, thanks to the built-in special-relativistic invariance. Secondly, there is no heating of masses, because the entangled tails represent virtual gravitons that scatter elastically. Thirdly, and most of all, by replacing the *corpuscules ultra-mondains* of Lesage by virtual gravitons – and thus by strands – we can predict an additional effect of gravity that is not described by the inverse square dependence: space-time curvature.

CURVED SPACE

In nature, observation shows that physical space is not flat around masses, i.e., in the presence of gravity. Near masses, physical space is *curved*. Observations also confirm that curved space-time remains 3+1-dimensional. The observation of curvature was predicted long before it was measured, because curvature follows unambiguously when the observer-invariance of the speed of light c and the observer-invariance of the gravitational constant G are combined.

We continue directly with the strand model of spatial curvature.

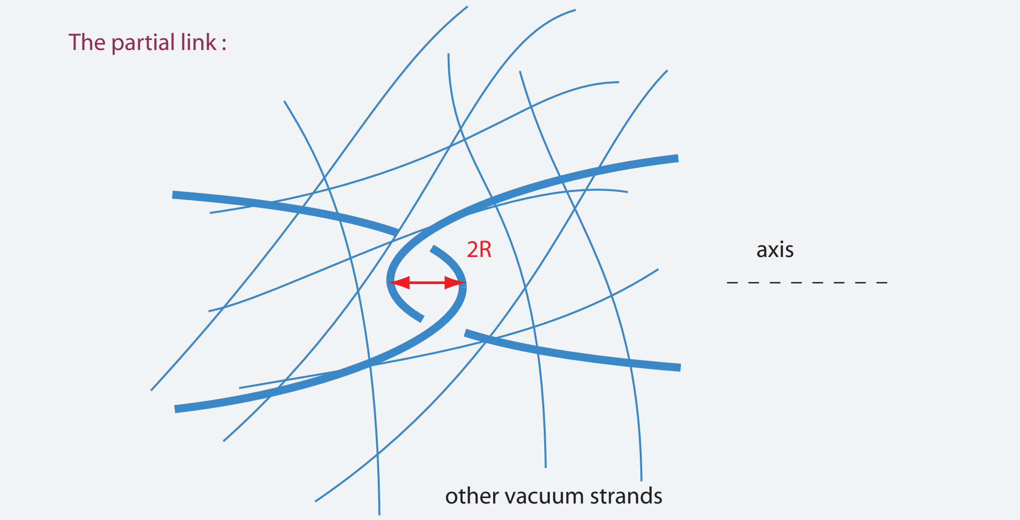


FIGURE 67 A schematic model of the fundamental defect, and thus the fundamental type of curvature: the *partial link*.

- ▷ In the case of curvature, *physical* space-time, which is due to averaged strand crossing switches, *differs* from *background* space-time, which usually corresponds to the tangent or asymptotic space-time. In Figure 67, the grey background colour can be taken as visualization of the background space.
- ▷ ***Curvature*** (of physical space-time) is due to simple, unknotted and weakly localized defects in the tangle of strands that make up the vacuum. An example is shown in Figure 67.
- ▷ ***Mass*** is a localized defect in space and is due to knotted or tangled strands. Thus mass curves space around it.
- ▷ ***Energy*** in a volume is the number of crossing switches per unit time. As a result, mass is equivalent to energy. As a second result, energy also curves space.
- ▷ ***Gravitation*** is the space-time curvature originating from compact regions with mass or energy.

These natural definitions show that curvature is due to strand configurations. In particular, curvature is built of unknotted – i.e., massless – *defects*. The massless defects leading to curvature are usually dynamic: they evolve and change. Such curvature defects – virtual gravitons – originate at regions containing matter or energy. In fact, the curvature of space around masses is a natural result of fluctuations of the strands that make up matter tangles. Page 306

We note that curved space, being a time average, is *continuous* and *unique*. Vacuum or curved space, more precisely, curved physical space, thus differs from background space, which is flat (and drawn in grey in the figures).

Incidentally, the distinction between physical and background space also avoids Ein-

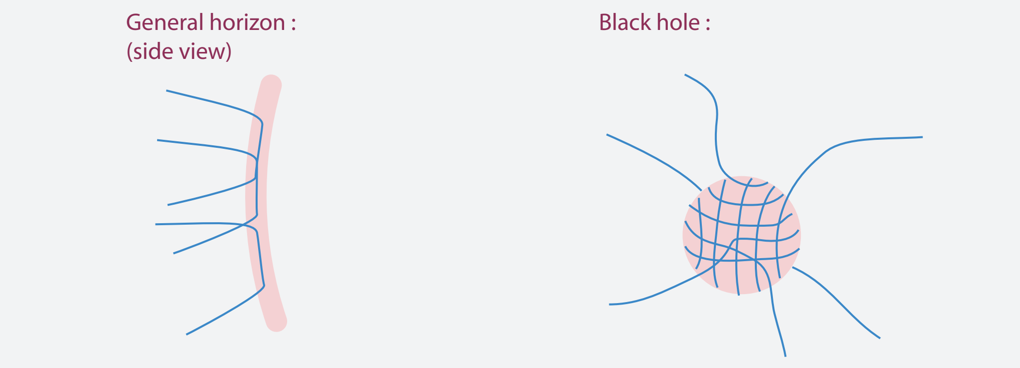


FIGURE 68 A schematic model of a general and a spherical horizon as tight weaves, as pictured by a distant observer. In the strand model there is *nothing*, no strands and thus no space, behind a horizon.

stein's hole argument; in fact, the distinction allows to discuss it clearly, since only physical space describes nature.

HORIZONS AND BLACK HOLES

In general relativity, another concept plays a fundamental role.

▷ A ***horizon*** is a tight, one-sided weave of strands.

Therefore, there are no strands behind the horizon. This implies that behind a horizon, there is no matter, no light, no space and no time – just *nothing*. Indeed, this is the experience of any observer about a horizon. A horizon is thus a structure that limits physical space, but not background space.

One particular type of horizon is well-known.

▷ A ***black hole*** is a tight, one-sided and *closed* weave of strands.

In principle, closed horizons can have any shape. The simplest case is the spherical, non-rotating horizon, which defines the *Schwarzschild black hole*. It is illustrated on the right-hand side of Figure 68.

If an observer is located outside a spherical horizon, the strand model states that there is nothing *inside* the horizon: no matter, no light and no vacuum. The strand model thus provides a simple and drastic view of black hole horizons. Figure 68 also illustrates that the concept of radius (or size) of a black hole has to be approached with the (well-known) care. In general, the size of a structure made of strands is the number of crossings encountered when travelling through it. However, an observer cannot travel *through* a black hole: there are no strands inside, thus there is no vacuum there! The size of a black hole must therefore be defined indirectly. The simplest way is to take the square root of the area, divided by 4π, as the radius. Thus the strand model, like general relativity, requires that the size of a compact horizon be defined by travelling *around* it.

IS THERE SOMETHING BEHIND A HORIZON?

A drawing such as the one of Figure 68 clearly points out the difference between the background space and the physical space. The *background space* is the space we need for thinking, and is the space in which the drawing is set. The *physical space* is the one that appears as a consequence of the averaging of the strand crossings. Physical space exists only outside the horizon. The physical space around a black hole is curved; it agrees with the background space only at infinite distance from the horizon. The strand model thus implies that there is *nothing*, not even a singularity, inside a black hole horizon.

Horizons are obviously observer-dependent. Both the existence and the shape of a horizon depends on the observer. As we will see, this happens in precisely the same way as in usual general relativity. In the strand model, there is no contradiction between the one observer who says that there is *nothing* behind a horizon, not even physical space, and another observer, who does not observe a horizon, and who says that there is *something* there. In the strand model, the two statements transform into each other under change of viewpoint. The transformation between the two viewpoints is a deformation of the involved strands.

We note that the equivalence of viewpoints and the statement that there is nothing behind a horizon is based on the combination of general relativity and quantum theory. If we would continue thinking that space and time is a manifold of points – thus disregarding quantum theory – these statements would *not* follow.

In summary, one-sided tight weaves are a *natural* definition of horizons.

ENERGY OF HORIZONS

The strand model allows to calculate the energy content of a closed horizon. Energy is action per unit time. In the strand model, the energy of a non-rotating spherical horizon is given by the number N_{cs} of crossing switches per time unit. In a tight weave, crossing switches cannot happen in parallel, but have to happen sequentially. As a result, a crossing switch 'propagates' to the neighbouring Planck area on the surface. Since the weave is tight and the propagation speed is one crossing per crossing switch time, this happens at the speed of light. In the time T that light takes to circumnavigate the sphere, all crossings switch. We thus have: Challenge 141 e

$$E = \frac{N_{cs}}{T} = \frac{4\pi R^2}{2\pi R}\,\frac{c^4}{4G} = R\,\frac{c^4}{2G}\,. \qquad (181)$$

Strands thus imply the well-known relation between energy (or mass) and radius of Schwarzschild black holes.

The tight-weave model of horizons also illustrates and confirms the *Penrose conjecture*. For a given mass, because of the minimum size of crossings, a spherical horizon has the smallest possible diameter, compared to other possible shapes. This implies that, for a given mass, spherical black holes are the densest objects in nature.

The strand model also naturally implies the *no-hair theorem*. Since all strands are the same, independently of the type of matter that fell into the horizon, a black hole has no characteristics other than mass, angular momentum and charge. The full argument is possible only after the end of the next chapter, when it will be clear that all particles

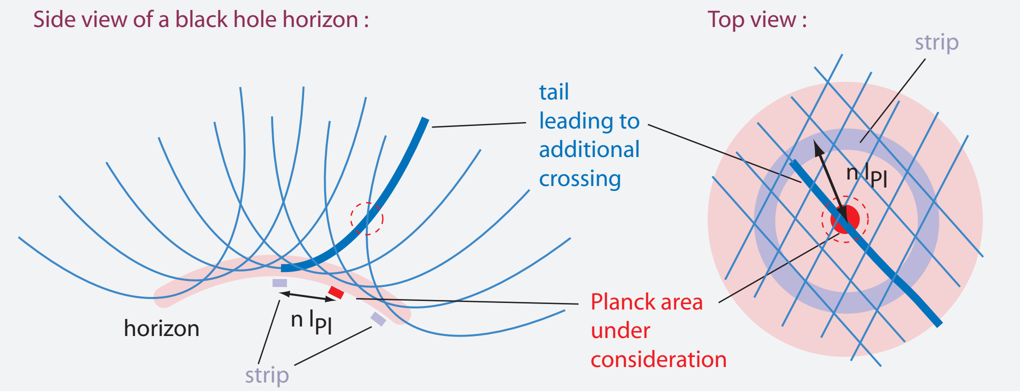


FIGURE 69 The entropy of black holes results from the number of possible crossing switches.

of the standard model are indeed made of the same featureless strands. But taking that result as given, the no-hair theorem follows. For example, strands explain naturally why neutral black holes made of antimatter and neutral black holes made of matter do not differ, if their masses and angular momenta are the same.

Entropy of horizons

Despite the tight weaving, the strands making up a horizon are still fluctuating and moving: the weave topology fluctuates. This fluctuating motion is the reason why horizons – in particular those of black holes – have entropy.

The entropy of a horizon is given by the natural logarithm of the number of its possible microstates times k. Here we again use the fundamental principle: a crossing switch defines the unit of entropy k. Obviously, the vacuum has vanishing entropy.

In the *absence of gravity*, the number of microstates of matter is determined as in usual thermodynamics (thermostatics), by the behaviour of tangle *cores*.

In *strong gravity*, when the distinction between matter and space is not so clear-cut, the number of microstates is determined by the possible crossing changes of the strands. In strong gravity, only *tails* play a role. This can be seen most clearly in the case of black holes.

To a first approximation, on each (corrected) Planck area of the horizon, the strands can cross in two different ways. The number N of Planck areas is given by $N^2 = Ac^3/4G\hbar$. The resulting number of microstates is 2^{N^2}. As mentioned, in the strand model, the entropy is given by the natural logarithm of the number of the possible microstates times k. This gives an entropy of a horizon of

$$S = A\,\frac{kc^3}{4G\hbar}\ln 2\ . \qquad (182)$$

This is the well-known first approximation of black hole entropy by 't Hooft: one bit per corrected Planck area. In the strand model, the proportionality of entropy and area is thus a direct consequence of the *extension* of the strands. This proportionality is also well known from studies of quantum gravity and of strings. In those approaches however,

Ref. 188

the relation between extension and the area proportionality is less obvious.

However, for Schwarzschild black holes, the entropy value of (182) is not correct. In the strand model, this incorrect value is explained as a consequence of neglecting the effects of the strand tails. Indeed, additional contributions to the entropy appear at a *finite distance* from the horizon, due to the crossing of the tails on their way to the border of space, as shown in Figure 69. The actual entropy will thus be larger than the first approximation, but still be proportional to the area A.

The correct proportionality factor between the area and the entropy of a black hole results when the strand tails are taken into account. (The correction factor is called the *Barbero–Immirzi parameter* in the literature on quantum gravity.) The calculation is simplest for Schwarzschild black holes. By construction, a black hole with macroscopic radius R, being a tight weave, has R/l_{Pl} tails. For each given Planck area, there are, apart from the basic, or lowest crossing, additional crossings 'above it', along the radial direction. These additional crossings are due to the tails from neighbouring and distant Planck areas, and are shown in Figure 69.

Taking into effect all strand tails allows us to calculate the average number of crossings *above* a given Planck area. The main point is to perform this calculation for all those tails that start in a circular strip of Planck width centred around the Planck area under consideration. We then add the probabilities for all possible circular strips. One such circular strip is drawn in Figure 69.

The definition of horizons as tight weaves implies that a horizon with N^2 Planck areas is made of N strands. This means that for each circular strip of radius nl_{Pl}, there is only *one* strand that starts there and reaches spatial infinity as a tail.

For this tail, the average probability p that it crosses above the central Planck area under consideration is Challenge 142 e

$$p = \frac{1}{n!} \,. \tag{183}$$

Summing over all strips, i.e., over all values n, we get a total of $\sum_{n=0}^{\infty} 1/n! = \mathrm{e} = 2.71828...$ microstates on and above the central Planck area under consideration. Thus the number e replaces the number 2 of the first approximation by 't Hooft. In other words, the number of horizon microstates of a Schwarzschild black hole is not 2^{N^2}, but e^{N^2}. As a consequence, the entropy of a macroscopic Schwarzschild horizon becomes

$$S = A\,\frac{kc^3}{4G\hbar} \,. \tag{184}$$

This is the Bekenstein–Hawking expression for the entropy of Schwarzschild black holes. The strand model thus reproduces this well-known result. With this explanation of the difference between 2 and e = 2.71828..., the strand model confirms the old idea that the entropy of a black hole is mainly located *at and near* the horizon.

The above calculation, however, counts some states more than once. Topologically identical spherical horizons can differ in the direction of their north pole and in their state of rotation around the north–south axis. If a spherical horizon is made of $2N$ strands, it has N^2 possible orientations for the north pole and N possible angular orientations around the north–south axis. The actual number of microstates is thus e^{N^2}/N^3.

Using the relation between N^2 and the surface area A, namely $A = N^2 4G\hbar/c^3$, we get the final result

$$S = A\,\frac{kc^3}{4G\hbar} - \frac{3k}{2}\ln\frac{A\,c^3}{4G\hbar}\ . \tag{185}$$

The strand model thus makes a specific prediction for the logarithmic correction of the entropy of a Schwarzschild black hole. This final prediction of the strand model agrees with many (but not all) calculations using strings or other quantum gravity approaches. Ref. 189

Temperature, radiation and evaporation of black holes

The strands that make up a horizon fluctuate in shape. Since every horizon contains energy, the shape fluctuations imply energy fluctuations. In other words, horizons are predicted to have a *temperature*. The value of the temperature can be deduced from the strand model by noting that the characteristic size of the fluctuations for a spherical horizon is the radius R of the horizon. Therefore we have

$$kT = \frac{\hbar c}{2\pi R}\ . \tag{186}$$

Using the definition of *surface gravity* as $a = c^2/R$, we get

$$T = \frac{\hbar a}{2\pi k c}\ . \tag{187}$$

The strand model predicts that horizons have a temperature proportional to their surface gravity. This result has been known since 1973. Ref. 56, Ref. 57

All hot bodies radiate. The strand model thus predicts that Schwarzschild black holes *radiate* thermal radiation of the horizon temperature, with power and wavelength

$$P = 2\pi\hbar c^2/R^2 \quad , \quad \lambda \approx R\ . \tag{188}$$

This is a well-known consequence of the temperature of black holes.

As thermodynamic systems, horizons follow thermodynamics. The strand model implies that in black hole radiation, there is *no* information loss. In the strand model, black hole radiation and evaporation occur by reduction of the number of strands that make up the horizon. The strand model thus predicts that black holes *evaporate completely*, until only elementary particles are left over.

Black hole limits

In many ways, black holes are extreme physical systems. Indeed, black holes realize many limits of nature; in this they resemble light, which realizes the speed limit. We now explore some of them.

For a general physical system, not necessarily bound by a horizon, the definitions of energy and entropy with strands allow some interesting conclusions. The entropy of a system is the result of the number of crossing possibilities. The energy of a system is the number of crossing changes per unit time. A large entropy is thus only possible if a system shows many crossing changes per time. Since the typical system time is given by the circumference of the system, the entropy of a physical system is therefore limited:

$$S \leqslant ER\, 2\pi k/\hbar c\ . \tag{189}$$

This relation is known as *Bekenstein's entropy bound*; it thus also follows from the strand model. The equality is realized only for black holes.

Horizons are the limit systems of general relativity. In the strand model, horizons are tight, one-sided weaves. For example, this implies that any tangle that encounters a horizon is essentially flat. Because of tangle flatness, at most one Planck mass can cross a horizon during a Planck time. This yields the mass rate limit $\mathrm{d}m/\mathrm{d}t \leqslant c^3/4G$ that is valid in general relativity and in nature.

Black holes can rotate. The strand model states that there is a highest angular frequency possible; it appears when the equator of the black hole rotates with the speed of light. As a result, the angular momentum J of a black hole is limited by $J < 2GM^2/c$. This limit is well known from general relativity. Ref. 58

In the strand model, a horizon is a tight weave. Therefore, a horizon cannot contain more elementary charges than it can contain knots. Thus, a horizon cannot contain more elementary charges than crossings. As a result, the strand model predicts that the maximum charge of a horizon is limited by its area. In other words, the charge limit of a non-rotating black hole is proportional to its mass.

For non-rotating black holes, the precise charge limit can be deduced rapidly. The force limit in nature implies that the electrical forces between two charged black holes must be lower than their gravitational interaction. This means that

$$\frac{Q^2}{4\pi\varepsilon_0 r^2} \leqslant \frac{GM^2}{r^2}\ , \tag{190}$$

or

$$Q^2 \leqslant 4\pi\varepsilon_0 GM^2\ . \tag{191}$$

This is the well-known charge limit for (static) black holes given by the Reissner–Nordström metric. It follows directly from the strand model.

The strand model limits energy density to the Planck energy per Planck volume, or $c^7/(16G^2\hbar)$. This implies that the strand model does not allow singularities, be they dressed or naked. Indeed, no singularity has ever been observed.

Of all non-rotating horizons, spherical horizons stand out. The strand model explains and visualizes all their properties. In fact, strands also illustrate the non-existence of (uncharged) one-dimensional or toroidal horizons in 3+1 space-time dimensions. Such configurations are unstable, in particular against transverse shear and rearrangement of the strands. The strand model thus confirms that spherical horizons are the most compact bodies with a given mass.

In summary, the strand model reproduces the known limit properties of horizons. And all these results are independent of the precise fluctuation details of the strands.

Curvature around black holes

The tails of a black hole extend up to the border of space; the density of tails is highest at the horizon. A black hole is therefore surrounded by partial links at any *finite* distance from the horizon. In other words, the space around a black hole is *curved*. The value of the space-time curvature increases as one approaches the horizon, because of the way in which the partial links hinder each other in their motion. The nearer they are to the horizon, the more they hinder each other. The curvature that appears is proportional to the density of partial links and to their average strand curvature.

At the horizon, the curvature radius is the horizon radius. By construction, the number of tails departing from a non-rotating black hole is proportional to the horizon radius R. Hence at a radial distance r from a static black hole, the spatial curvature K is

$$K \sim \frac{R}{r^3} \,. \tag{192}$$

So at the horizon itself, the curvature K is (of the order of) the inverse square of the horizon radius; further away, it decreases rapidly, with the third power of the distance. This result is a well-known property of the Schwarzschild solution. The rapid decay with radius is the reason why in everyday situations there is no noticeable curvature of space-time. In short, strands allow us to deduce the correct curvature of space-time around spherical masses.

In the strand model, it is easy to deduce that non-rotating horizons tend to be spherical: spheres are the bodies with the smallest surface for a given volume. The minimum surface appears because the strands, through their fluctuations, effectively 'pull' on each Planck area of the horizon. As a result, non-rotating macroscopic horizons will be spherical. (Deviations from the spherical shape will mainly occur at around the Planck scale.) With the definition of gravity waves given below, it also becomes clear that strongly deformed, macroscopic and non-spherical horizons are unstable against emission of gravity waves or particles.

In summary, strands reproduce all known qualitative and quantitative properties of horizons and of black holes, and thus of general systems with strong gravitational fields. All predictions from strands agree with observations and other approaches to quantum gravity. These are the first hints that strands imply the field equations.

The field equations of general relativity

The field equations can be deduced from the fundamental principle in two different, but related ways. Essentially, both derivations repeat the reasoning for classical gravitation given above. The first deduction of the field equations is based on an old argument on the thermodynamics of space-time. Strands show that horizons obey

Page 249

Ref. 19

- an area–entropy relation of $S = A\,kc^3/4G\hbar$,
- a curvature–temperature relation of $T = a\,\hbar/2\pi kc$,
- a relation between heat and entropy of $\delta Q = T\delta S$.

Using these three properties, and using the relation $\delta Q = \delta E$ – valid *only* in case of horizons – we get the first law of horizon mechanics

$$\delta E = \frac{c^2}{8\pi G} a\, \delta A\ . \tag{193}$$

Page 29 From this relation, using the Raychaudhuri equation, we obtain the field equations of general relativity. This deduction was explained earlier on.*

In other words, the field equations result from the thermodynamics of strands. It is worth noting that the result is independent of the details of the fluctuations, as long as the three thermodynamical properties are valid. We can turn this argument around. Strand fluctuations *must* obey these properties to allow to define space-time. If they obey these properties, then space-time exists and curves according to general relativity.

The second derivation of the field equations of general relativity follows the spirit of the strand model most closely. Strands imply that all physical quantities are limited by the corresponding Planck limit. These limits are due to the limit to the fundamental principle, in other words, they are due to the packing limit of strands. In particular, the Ref. 16 fundamental principle limits force by $F \leqslant c^4/4G$ and power by $P \leqslant c^5/4G$. We have Page 29 shown above that this limit implies the field equation.

In summary, the strand model asserts that the field equations appear as consequences of fluctuations of featureless strands that cannot cross. The strand model thus implies that a horizon and a particle gas at Planck energy do not differ. However, the value of the cosmological constant is not predicted from strand thermodynamics.

EQUATIONS FROM NO EQUATION

The strand model asserts that the field equations are not the result of another, more basic evolution equation, but result directly from the fundamental principle. To say it bluntly, the field equations are deduced from a drawing – the fundamental principle shown in Page 149 Figure 12. This strong statement is due to a specific property of the field equations and to two properties of the strand model.

---

* Here is the argument in a few lines. The first law of horizon mechanics can be rewritten, using the energy–momentum tensor T_{ab}, as

$$\int T_{ab} k^a \mathrm{d}\Sigma^b = \frac{c^2}{8\pi G} a\, \delta A$$

where $\mathrm{d}\Sigma^b$ is the general surface element and k is the Killing vector that generates the horizon. The Raychaudhuri equation allows to rewrite the right hand side as

$$\int T_{ab} k^a \mathrm{d}\Sigma^b = \frac{c^4}{8\pi G} \int R_{ab} k^a \mathrm{d}\Sigma^b$$

where R_{ab} is the Ricci tensor describing space-time curvature. This equality implies that

$$T_{ab} = \frac{c^4}{8\pi G} (R_{ab} - (R/2 + \Lambda) g_{ab})$$

where Λ is an undetermined constant of integration. These are Einstein's field equations of general relativity. The field equations are valid everywhere and for all times, because a coordinate transformation can put a horizon at any point and at any time.

First of all, the field equations are essentially consequences of the thermodynamics of space-time. In the strand model, the thermodynamic properties are deduced as a consequence of the strands. This deduction does not require underlying evolution equations; the result follows from the statistical behaviour of strands.

An essential property of the model is its independence from the underlying motion of the strands. In the strand model one gets evolution equations – the field equations in this case – without deducing them from another equation. The deduction from the strand model works for *any* underlying evolution equation, as long as the thermodynamic properties of the strand fluctuations are reproduced.

The last property that allows to deduce the field equations directly from a graph, and not from another equation, is the relation between the graph and natural physical units. The relation with natural units, in particular with the quantum of action $\hbar$ and the Boltzmann constant k, is fundamental for the success of the strand model.

The discussion so far adds another aspect: unique, underlying, more basic evolution equations *cannot* exist. There are two reasons. First, a unique underlying equation would itself require a deduction, thus would not be a satisfying solution to unification. Secondly, and more importantly, evolution equations are differential equations; they assume well-behaved, smooth space-time. At Planck scales, this is impossible. Any principle that allows to deduce the field equations cannot itself be an evolution equation.

THE HILBERT ACTION OF GENERAL RELATIVITY

We have just shown that the strand model implies the field equations of general relativity. Page 194 We have also shown above that, in the strand model, the least action principle is a natural property of all motion of strands. Combining these two results, we find that a natural way to describe the motion of space-time is the (extended) *Hilbert action* given by

$$W = \frac{c^3}{16\pi G} \int (R - 2\Lambda)\, \mathrm{d}V \quad , \qquad (194)$$

where R is the Ricci scalar, $\mathrm{d}V = \sqrt{\det g}\, \mathrm{d}^4 x$ is the invariant 4-volume element of the metric g, and Λ is the cosmological constant. As is well known, the description of evolution with the help of an action does not add anything to the field equations; both descriptions are equivalent.

GRAVITONS AND GRAVITATIONAL WAVES

In the strand model, gravitons can be seen as a special kind of partial links. An example is shown in Figure 70. As a twisted pair of parallel strands, the graviton returns to itself after rotation by π; it thus behaves like a spin-2 boson, as required.

Can single gravitons be observed? The strand model implies that the absorption of a single graviton by an elementary particle changes its spin or position. However, such a change cannot be distinguished by a quantum fluctuation. The strand model also predicts that gravitons do not interact with photons, because they have no electric charge. In summary, the strand model predicts that single gravitons *cannot* be detected.

The situation changes for gravitational waves. Such waves are coherent superpositions

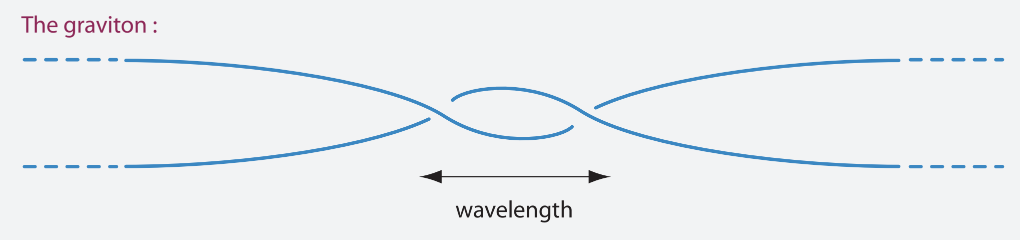


FIGURE 70 The graviton in the strand model.

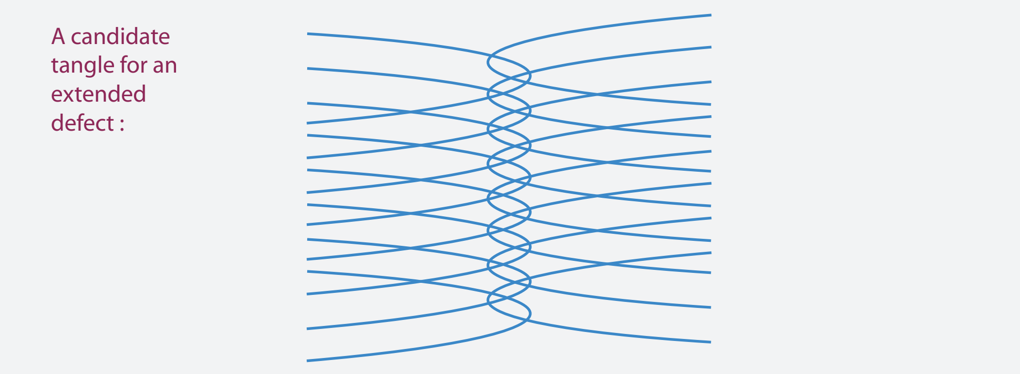


FIGURE 71 A speculative, highly schematic model for a cosmic string, a one-dimensional defect in space-time

of large numbers of gravitons. In such a case, the argument against the detection of single gravitons does not apply. In short, the strand model predicts that gravitational waves *can* be observed.

OPEN CHALLENGE: IMPROVE THE ARGUMENT FOR THE GRAVITON TANGLE

The argument that leads to the graviton tangle is too hand-waving. Can you make the argument more compelling? Challenge 143 ny

OTHER DEFECTS IN VACUUM

The strand model provides a quantum description of gravitation. The strand model does so by defining physical space as the average of the crossing switches induced by strand fluctuations among untangled strands. Matter, radiation and horizons are defects in these untangles strands.

So far, we have been concerned with localized, i.e., zero-dimensional, defects and with horizons, i.e., two-dimensional defects. But modelling of the vacuum as a set of untangled strands also implies the possible existence of one-dimensional defects that are equivalent to dislocations and disclinations in solids, of other two-dimensional defects, or of three-dimensional defects. Such defects could model cosmic strings, domain walls, wormholes, toroidal black holes, time-like loops and regions of negative energy.

A simple example of a possible defect is shown in Figure 71. The illustration can be seen as the image of a one-dimensional defect or as the cross section of a two-

TABLE 11 Correspondences between physical systems and mathematical tangles.

PHYSICAL SYSTEM	STRANDS	TANGLE TYPE
Vacuum	many infinite unknotted strands	unlinked
Dark energy	many fluctuating infinite strands	unlinked
Elementary vector boson	one infinite strand	knotted or unknotted curve
Quark	two infinite strands	rational tangle
Lepton	three infinite strands	braided tangle
Meson, baryon	three or more infinite strands	rational tangle
Higher-order propagating fermion	two or more infinite strands	locally knotted or prime tangle
Virtual particles	open, unlinked and closed strands	trivial tangles, knots, links
Composed systems	many strands	separable tangles
Graviton	two infinite twisted strands	specific rational tangle
Gravity wave	many infinite twisted strands	many graviton tangles
Horizon	many tightly woven infinite strands	web-like tangle

dimensional defect. Are such defects stable against fluctuations? The strand model suggests that they are not; they should decay into a mixture of gravitons, black holes, matter and radiation particles. However, this issue is still a topic of research, and will not be covered here.

Exploring the stability of wormholes, time-like loops and toroidal black holes leads to similar results. It seems that the strand model should not allow time-like loops of macroscopic size, since any place that cannot be embedded locally into three flat spatial dimensions is either a particle or a black hole. Alternatively, macroscopic time-like loops would collapse or decay because of the fluctuations of the strands. In the same way, wormholes or black holes with non-trivial topology should be unstable against more usual strand structures, such as particles or black holes.

We also note the strand model does not allow volume defects (black holes being surface-like defects). The most discussed types of volume defect are macroscopic regions of negative energy. Energy being action per unit time, and action being connected to crossing changes, the model does not allow the construction of negative-energy regions. However, the strand model does allow the construction of regions with lower energy than their environment, as in the Casimir effect, by placing restrictions on the wavelengths of photons.

The final and general connection between tangle types and defects is shown (again) in Table 11. The next chapter will give details of the tangles corresponding to each particle.

In summary, the strand model reproduces the results of modern quantum gravity and predicts that none of the more spectacular defects conjectured in the past – linear defects such as cosmic strings, surface defects such as wormholes, volume defects such as negative-energy regions – can appear in nature.

Torsion, curiosities and challenges about general relativity

On one hand, the strand model denies the existence of any specific effects of *torsion* on gravitation. Ref. 190 On the other hand, the strand model of matter describes spin with the belt trick. The belt trick is thus the strand phenomenon that is closest to the idea of torsion. Therefore, exaggerating a bit in the other direction, it can also be argued that in the strand model, torsion effects are quantum field theory effects.

* *

The strand model describes three-dimensional space as made of tangled strands. Several similar models have also been proposed. Above all, the model of space as a *nematic world crystal* stands out as the most similar. This model was proposed by Hagen Kleinert in Ref. 191 the 1980s. It took its inspiration from the famous analogy by Ekkehart Kröner between the equations of solid-state elasticity around line defects and the equations of general Ref. 192 relativity.

Ref. 158 Also the mentioned posets have been proposed as the fundamental structure of space in the 1980s.

Ref. 193 Various models of loop quantum gravity from the 1990s, inspired by spin networks, spin foams and by similar systems, model empty space as made of extended constituents. These extended constituents tangle, or bifurcate, or are connected, or sometimes all of this at the same time. Depending on the model, the constituents are lines, circles or ribbons. In some models their shapes fluctuate, in others they don't.

Ref. 159 Another type of Planck-scale crystal model of the vacuum has been proposed by David Finkelstein. In 2008, a specific model of space, a crystal-like network of connected Ref. 194 bifurcating lines, has been proposed by Gerard 't Hooft.

All these models describe space as made of some kind of extended constituents in a three-dimensional background. All these models derive general relativity from these constituents by some averaging procedure. The lesson is clear: it is not difficult to derive general relativity from a microscopic model of space. As Luca Bombelli said already in the early 1990s, the challenge for a microscopic model is not to derive general relativity; the real challenge is to derive the other interactions. So far, the strand model seems to be the only model that has proposed a solution.

* *

In September 2010, two years after the strand model appeared, the model starts being confirmed by independent research. In a extended article exploring the small scale structure of space-time from various different research perspectives in general relativity, Steven Carlip comes to the conclusion that all these perspectives suggest the common idea that Ref. 156 "space at a fixed time is thus threaded by rapidly fluctuating lines".

* *

The strand model assumes that space is not defined at the cosmic horizon, and that therefore, strand impenetrability does not hold there. Does the same occur at a black hole horizon?

* *

Page 32 The strand model also allows to answer the question whether quantum particles are black holes: no, they are not. Quantum particles are tangles, but they do not have horizons.

* *

Challenge 144 ny Can black hole radiation be seen as the result of trying to tear vacuum apart?

* *

The strand model makes the point that *entanglement* and empty space – and thus quantum gravity – have the same nature: both are due to crossing strands. This idea has been explored independently by Mark van Raamsdonk. Ref. 195

* *

Argue or show that the proportionality of entropy and area implies that black holes are made of strands. Challenge 145 ny

* *

Argue or show that no thermodynamic system that is *not* equivalent to strands can reproduce general relativity. Challenge 146 ny

PREDICTIONS OF THE STRAND MODEL ABOUT GENERAL RELATIVITY

As just presented, the strand model makes several verifiable predictions.

- The maximum energy speed in nature is c, at all energy scales, in all directions, at all times, at all positions, for every physical observer. This agrees with observations.
- No deviations from special relativity appear for any measurable energy scale, as long as gravity plays no role. No 'Double' or 'Deformed Special Relativity' holds in nature, even though a maximum energy-momentum for elementary particles does exist in nature. This agrees with observations.
- There is a maximum power or luminosity $c^5/4G$, a maximum force or momentum flow $c^4/4G$, and a maximum mass change rate $c^3/4G$ in nature. This agrees with observations, but experimental data is far from these limit values.
- There is a minimum distance and a minimum time interval in nature. There is a maximum curvature and a maximum mass density in nature. There are no singularities in nature. All this agrees with observations, but experimental data is far from sufficient.
- There are no deviations from general relativity, as described by the Hilbert action, for all measurable scales. The only deviations appear in situations with a few strands, i.e., in situations where quantum theory is necessary. This agrees with observations, but experimental data is far from sufficient.
- The usual black hole entropy given by Bekenstein and Hawking holds. The value has never been measured, but is consistently found in all calculations performed so far.
- There is no modified Newtonian dynamics, or MOND, with evolution equations that differ from general relativity. This agrees with most recent observations on galaxies, but experimental data is not sufficient.
- There is no effect of torsion that modifies general relativity. This agrees with observations.

- There is no effect of higher derivatives of the metric on the motion of bodies. This agrees with observations, but experimental data is far from sufficient.
- Observations are independent of the precise strand fluctuations. Mathematical consistency checks are possible.
- No wormholes, no negative energy regions and no time-like loops exist. This agrees with observations, but experimental data is far from complete.
- The Penrose conjecture holds. Here, a mathematical consistency check is possible.
- There are no cosmic strings and no domain walls. This agrees with observations, but experimental data is far from sufficient.
- Gravitons have spin 2; they return to their original state after a rotation by π and are bosons. This agrees with expectations.
- Gravitational waves exist and can be detected. This agrees with known data, but direct detection is still missing.

All listed predictions are unspectacular; they are made also by other approaches that contain general relativity and quantum gravity as limiting cases. In particular, the strand model, like many other approaches, predicts that, with the exception of the cosmological constant, *no quantum gravity effects will be observed*. Gravity will never yield new measurable quantum effects.

In other words, we have found *no unexpected* experimental predictions from the strand model in the domain of quantum gravity. But this is not a surprise. There are only two domains in which we can expect surprising predictions: cosmology (including the value of the cosmological constant) and particle physics. The rest of this chapter deals with cosmology. The subsequent chapter focuses on particle physics. Ref. 96

COSMOLOGY

Cosmology is an active field of research, and new data is collected all the time. We only give a short summary.

The sky is dark at night. This and other observations show that the universe is surrounded by a horizon and is of finite size and age. Precise measurements show that cosmic age is around 13 700 million years. The universe expands; the expansion is described by the field equations of general relativity. The universe's expansion accelerates; the acceleration is described by the *cosmological constant*, the so-called *dark energy*, that is small, but non-vanishing and of positive value. The universe is observed to be flat, and, averaged over large scales, homogeneous and isotropic. The observed average matter density in the universe is about 18 times smaller than the energy density due to the cosmological constant. In addition, there is a large amount of matter around galaxies that does not radiate; the nature of this *dark matter* is unclear. Galaxy formation started from early density fluctuations; the typical size and amplitude of the fluctuations are known. The topology of space is observed to be simple.

The strand model, like any unified description of nature, must reproduce and explain these measurement results. Otherwise, the strand model is wrong.

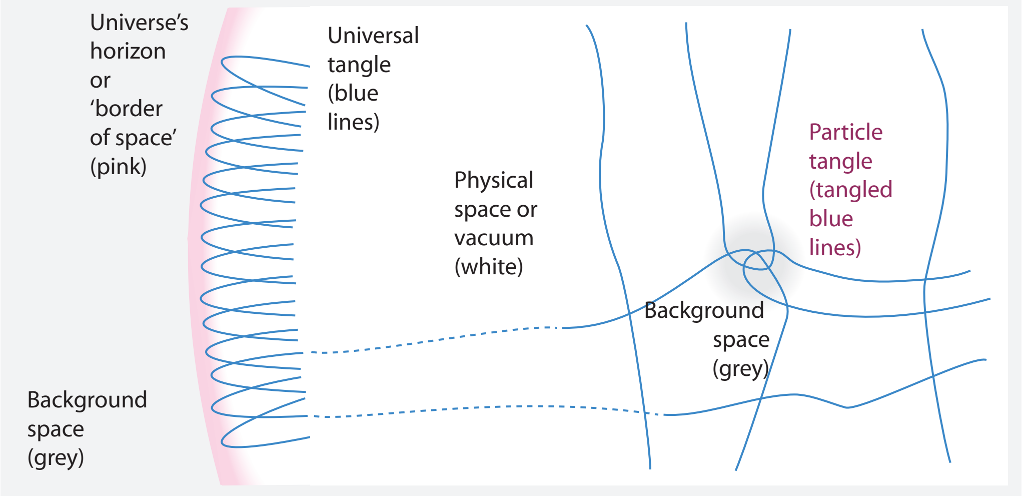


FIGURE 72 In the strand model, the universe is limited by a horizon, as schematically illustrated here. Physical space (white) matches background space (grey) only inside the horizon. Physical space only exists inside the cosmic horizon.

The finiteness of the universe

In the strand model, cosmology is based on one idea:

▷ The ***universe*** is made of *one* fluctuating strand. Fluctuations increase the complexity of the strand knottedness over time.

The existence of finite size and of finite age then follows automatically:

▷ The ***universe's horizon*** appears at the age or distance at which the strand crossings cannot be embedded any more into a three-dimensional background space.

The strand model thus has a simple explanation for the finiteness of the universe and the horizon that bounds it. A schematic illustration is given in Figure 72.

The strand model predicts that the cosmic horizon is an *event horizon*, like that of a black hole. Until 1998, this possibility seemed ruled out by experiment; but in 1998, it was discovered that the expansion of the universe is accelerating. This discovery implies that the cosmic horizon is indeed an event horizon, as predicted by the strand model. In fact, the strand model predicts that all horizons in nature are of the same type. this also means that the universe is predicted to saturate Bekenstein's entropy bound. In fact, the strand model predicts that the universe is a kind of *inverted back hole*. Like for any situation that involves a horizon, the strand model thus does not allow to make statements about properties 'before' the big bang or 'outside' the horizon. *

Ref. 196
Ref. 197

---

* In particular, the strand model states that the matter that appears at the horizon during the evolution of the universe appears through Bekenstein–Hawking radiation. This contrasts with the 'classical' explanation that it simply crosses the horizon.

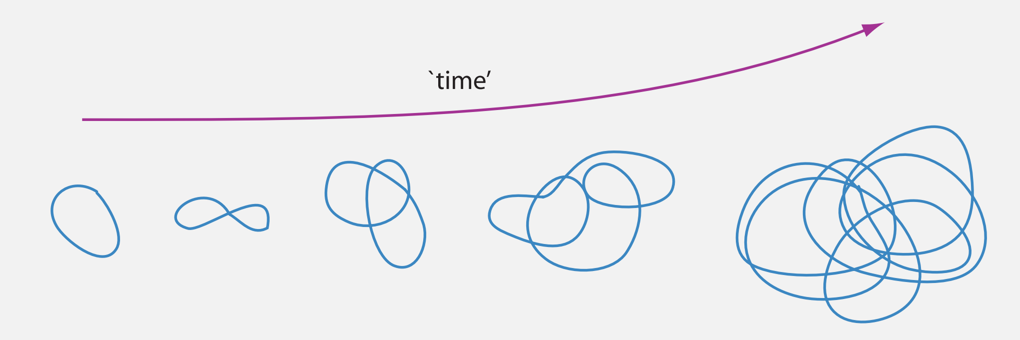


FIGURE 73 An extremely simplified view of how the universe evolved near the big bang. In this situation, physical space is not ‘yet’ defined.

We note that modelling the universe as a single strand implies that it contains tangles. In other words, the strand model makes the prediction that the universe cannot be empty, but that it must contain particles.

We also note that describing the universe as made of a single strand is a natural, but somewhat unusual way to incorporate what particle physicists and cosmologists call *holography*. As a consequence, strand cosmology naturally reproduces holographic cosmology.

> “ Or cette liaison ou cet accommodement de toutes les choses créées à chacune, et de chacune à toutes les autres, fait que chaque substance simple a des rapports qui expriment toutes les autres, et qu’elle est par conséquent un miroir vivant perpétuel de l’univers.* ”
>
> Gottfried Wilhelm Leibniz, *Monadologie*, 56.

The big bang

Ref. 198 Page 258 Any expanding, homogeneous and isotropic matter distribution had earlier stages of smaller size and higher density. But the strand model also states that singularities do not appear in nature, because there is a highest possible energy density. This implies that the big bang might be imagined as illustrated in Figure 73. Obviously, physical space and time are not well defined near that situation, so that the figure has to be taken with a grain of salt. Nevertheless, it shows how the evolution of the universe can be seen as resulting from the increase in tangledness of the strand that makes up nature.

It is expected that the evolution of the strand model just after the big bang automatically leads to a homogeneous and isotropic matter distribution and to flat space. Also the scale invariance of early density fluctuations seems natural in the strand model. In short, the strand model looks like a promising alternative to *inflation*: the hypothesis of inflation becomes unnecessary in the strand model, because the strand model makes the same predictions already. This issue is still subject of research.

---

* ‘Now this connexion or adaptation of all created things to each and of each to all, means that each simple substance has relations which express all the others, and, consequently, that it is a perpetual living mirror of the universe.’

The cosmological constant

In the strand model, *vacuum energy*, or *dark energy*, is due to the cosmological constant, which itself is due to strand fluctuations. As we saw above, the strand model predicts that the cosmological constant Λ for infinitely extended flat space vanishes, because the energy density vanishes in that case. But the strand model also predicts that for *finite* extension, the cosmological constant does *not* vanish. Indeed, in the strand model, a finite size limits the fluctuations of the strands. Fluctuations with sizes larger than the size of space are frozen out; this leads to an effective repulsion of strands. This leads to a cosmological constant given by (the square of) the extension of space: $\Lambda = 1/R^2_{\text{max}}$. In particular, the strand model predicts a small *positive* cosmological constant, i.e., a constant that leads to a small repulsion of masses. Page 196

The relation between the cosmological constant und the radius can be found also with another, more precise argument, based on holography, and given by Balázs and Szapudi. Ref. 200 Bekenstein's holographic entropy bound states that for all systems of size R and energy E one has

$$S \leqslant ER\,\frac{2\pi k}{\hbar c}\;. \qquad (195)$$

For a spherical system, this yields

$$S \leqslant A\,\frac{kc^3}{4G\hbar}\;. \qquad (196)$$

The application of this inequality to the universe is the Fischler–Susskind holographic conjecture. Ref. 199 Using the energy–entropy relation $E = TS$ valid for any holographic system, and introducing the *energy* density ρ_{E}, we get the limit given by

$$\rho_{\text{E}} \leqslant \frac{T}{R}\,\frac{3kc^3}{4\hbar G}\;. \qquad (197)$$

Using the formula for temperature $T = \hbar c/2\pi kR$ for a horizon found by Gibbons, Bekenstein and Hawking, we get

$$\rho_{\text{E}} \leqslant \frac{1}{A}\,\frac{3c^4}{2G} = \frac{1}{4\pi R^2}\,\frac{3c^4}{2G} \qquad (198)$$

The strand model predicts that the universe *saturates* the entropy bound. In other words, assuming that R is c times the age of the universe t_0, the strand model predicts that the total energy density of the universe is equal to the so-called *critical* energy density.

The equality of the measured total energy density and the critical density is well known. Ref. 201 These measurements show that the present total energy density of the universe is about

$$\rho_{\text{E vac}} \approx 8.5 \cdot 10^{-10}\,\text{J/m}^3 \quad \text{or} \quad \rho_{\text{m vac}} = 0.94(9) \cdot 10^{-26}\,\text{kg/m}^3)\;. \qquad (199)$$

In other words, the strand model, like the holographic argument, predicts that the cosmological constant is limited by

$$\Lambda \leqslant \frac{3}{c^2 t_0^2} \ . \tag{200}$$

Ref. 201 Modern measurements yield 74% of the maximum possible value.

The argument for the cosmological constant can be made for any age of the universe. Therefore, the strand model predicts that the cosmological constant Λ *decreases* with increasing radius of the universe. Ref. 202 In particular, there is no need for a scalar field that makes the cosmological constant decrease; the decrease is a natural result of the strand model. The strand model states that the cosmological constant appears in the field equations as a quantum effect due to the finite size of the universe. The strand model thus implies that there is no separate equation of motion for the cosmological constant, but that the constant appears as a large-scale average of quantum effects, as long as the size of the universe is limited.

In other words, the strand model predicts that like the field equations of general relativity, also the expansion and the acceleration of the universe result from strand fluctuations. Ref. 203 In particular, the strand model implies that the effect recently proposed by Wiltshire – that the cosmological constant is an artefact the inhomogeneity of matter distribution – is *not* fundamental, but may at most influence the value somewhat. (Could the difference between the maximum possible and the measured value of the cosmological constant be due to this effect?) Challenge 147 ny

The value of the matter density

The strand model predicts that horizons emit particles. As a consequence, the strand model predicts an upper limit for the number N_b of baryons that could have been emitted by the cosmic horizon during its expansion. For a horizon shining throughout the age of the universe t_0 while emitting the maximum power $c^5/4G$, we get

$$N_{\text{b}0} \leqslant \frac{t_0 \ c^5/4G}{m_\text{b} c^2} = 2.6 \cdot 10^{79} \ . \tag{201}$$

Equality would hold only if the contributions of photons, electrons, neutrinos and dark matter could be neglected. In short, using the age $t_0 = 13.7$ Ga, the strand model predicts that at most $2.6 \cdot 10^{79}$ baryons exist in the universe at present. Modern measurements Ref. 201 indeed give values around this limit.

In other terms, the strand model states that the sum of all particle energies in the universe is at most $t_0 c^5/4G$, or 50% of the critical density; this includes observable as Ref. 201 well as dark matter. The experimental value is about 26% of the critical density. We will Page 305 discuss the nature of dark matter later on.

The strand model also makes a clear statement on the change of matter density with time. As just explained, the number of baryons is predicted to increase with time t, due to their appearance at the horizon. Also the radius will increase (roughly) with time; as a result, the strand model predicts that matter density decreases as $1/t^2$. This unexpected prediction contrasts with the usually assumed $1/t^3$ dependence in a matter-dominated universe. The prediction has yet to be tested with observations.

We note that these arguments imply that the ratio between matter density and vacuum energy density is a quantity related to the details of the radius increase during the history of the universe.

Open challenge: Are the dark energy and matter densities correct?

Challenge 148 ny

In the arguments above, is there a factor of 2 missing somewhere that induces incorrect conclusions about dark matter density? Might the prediction of dark matter increase, decrease or even disappear after correction of this missing numerical factor? How does this affect the galaxy rotation curves?

The topology of the universe

In the strand model, physical space-time, whenever it is defined, cannot be multiply connected. All quantum gravity approaches make this prediction, and the strand model confirms it: since physical space-time is a result of averaging strand crossing switches, non-trivial topologies (except black holes) do not occur as solutions. For example, the strand model predicts that wormholes do not exist. In regions where space-time is undefined – at and beyond horizons – it does not make sense to speak of space-time topology. In these regions, the fluctuations of the universal strand determine observations. In short, the strand model predicts that all searches for non-trivial *macroscopic* topologies of the universe, at both high and low energies, will yield negative results. So far, this prediction agrees with all observations.

Predictions of the strand model about cosmology

In the domain of cosmology, the strand model makes the following verifiable predictions.

- The universe is not empty. (Agrees with observation.)
- Its integrated luminosity saturates the power limit $c^5/4G$. (Agrees with observation.)
- The energy density of the universe saturates the entropy bound. (Agrees with observation.)
- The are no singularities in nature. (Agrees with observation.)
- Dark energy exists and results from vacuum/strand fluctuations. (Agrees with observation.) Dark energy, or vacuum energy, is due to the cosmological constant.
- The cosmological constant Λ is positive and changes with the radius R of the universe as $1/R^2$. (This prediction differs from the usual cosmological models, which assume that Λ is constant or changes with time in other ways. The strand prediction might be checked in the near future by testing whether the minimum acceleration around galaxies changes with distance – if this minimum is related to Λ.)
- The number of baryons in nature is limited by the maximum luminosity times the age of the universe. The present upper limit is $2.6 \cdot 10^{79}$ baryons. (Agrees with observation.)
- The matter density of the universe decreases with age as $1/t^2$. (Checks are under way. This prediction differs from the usual cosmological models.)
- There is nothing behind the cosmic horizon. Matter, energy and space appear at the horizon. (Agrees with observations and requirements of logic.)

- Early density fluctuations are scale-invariant. (Agrees with observation.)
- The universe is flat and homogeneous. (Agrees with observation.)
- Inflation is unnecessary. (Checks still need to be developed.)
- The universe's topology is trivial. (Agrees with observation.)
- The above statements are independent of the precise fluctuation details. (Can be tested with mathematical investigations.)

All these predictions can and will be tested in the coming years, either by observation or by computer calculations.

SUMMARY ON MILLENNIUM ISSUES: RELATIVITY

We have deduced special relativity, general relativity and cosmology from the strand model. In simple terms, space is the average of untangled strands. Space can be seen as a thermodynamic average due to strand shape fluctuations. The fundamental principle of the strand model implies the invariant Planck units and of the Lagrangian of general relativity.

The strand model explains the number of space-time dimensions, the vacuum energy density and the matter density of the universe. As shown in the next chapter, dark matter is predicted to be a combination of conventional matter and black holes. The initial condition issue has been defused. The cosmological constant has been related to the size of the universe. The topology of the universe has been clarified.

Page 305

The most important predictions of the strand model are the change of the cosmological constant with time and the absence of inflation. Various experiments will test these predictions with increased precision in the coming years. So far, measurements do not contradict these predictions.

The strand model confirms that the speed of light c and the corrected Planck force $c^4/4G$ are *limit* values. The strand model also predicts that no variation in space and time of c, G, $\hbar$ and k can be detected, because they define all measurement units.

The strand model also predicts that the cosmological constant and the masses of the elementary particles are the *only* quantum effects that will be observed in the study of gravitation.

Page 246

If we look at the millennium list of open issues in physics, we see that – except for the issue of dark matter – all issues about general relativity and cosmology have been settled. Above, we had already shown that the strand model explains all mathematical structures that appear in quantum theory and in particle physics. With the results from this chapter we can now say that the strand model explains *all* mathematical structures that appear in physical theories. In particular, strands explain the metric, curvature, wave functions and field intensities. All these quantities result from averages of crossing switches. But we are not done yet: we need to deduce the possible elementary particles and explain their properties.

Page 158
Page 245

Chapter 12

PARTICLES AND THEIR PROPERTIES DEDUCED FROM STRANDS

> No problem can withstand the assault of sustained thinking.
>
> Voltaire

Ref. 204

Strands describe quantum theory, gauge interactions and general relativity. But do trands also settle all issues left open by twentieth-century physics? Do they ettle the origin of all the elementary particles, their quantum numbers, their masses and their mixing angles? Do strands explain the coupling constants? In the millennium list of open issues in fundamental physics, these are the issues that remain. The strand model is correct only if these issues are resolved. Page 158

PARTICLES, QUANTUM NUMBERS AND TANGLES

In nature, we observe three entities: vacuum, horizons, and particles. Of these, *particles* are *localized* entities with specific *intrinsic* properties, i.e., properties that do not depend on their motion.

In nature, all intrinsic properties of all particles – in fact, those of every object and every image – are completely described by three *basic* properties: (1) the behaviour under space-time transformations, (2) the interactions and (3) the elementary particle content. The full list of these basic properties of particles is given in Table 12. Given the basic properties for each particle, and using the properties of the three gauge interactions, we can deduce *all* intrinsic particle properties that are *not* listed – such as half life, decay modes, branching ratios, electric dipole moment, T-parity, gyromagnetic ratio, electric polarizability etc. – and we can deduce all properties of objects and images – such as size, shape, colour, density, elasticity, brittleness, magnetism, conductance etc. Understanding *all* properties of matter and images thus only requires understanding the *basic* properties of the *elementary* particles.

The strand model states that all elementary (and all composed) particles are tangles of strands. This leads us to ask: Which tangle is associated to each elementary particle? What kinds of elementary particles are possible? Do these tangles reproduce, for each elementary particle, the observed values of the basic properties listed in Table 12? Page 166

It turns out that the strand model only allows a *limited number* of *elementary* particles. In addition, the tangles of these elementary particle have intrinsic properties that *match* the observed properties. To prove these strong statements, we explore tangles according

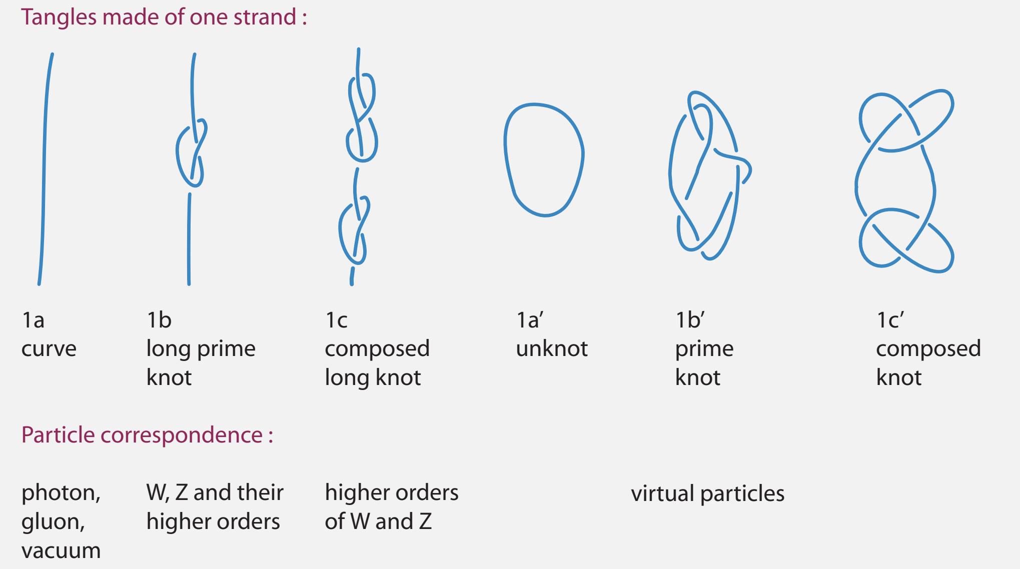


FIGURE 74 Examples for each class of tangles made of one strand.

to the number of strands they are made of.

PARTICLES MADE OF ONE STRAND

In the strand model, all particles made of *one* strand have spin 1, are elementary, and are bosons. Conversely, all elementary particles of spin 1 can only have two tails, and thus must be made of a single strand. Only one-stranded tangles return to the original strand after core rotation by 2π. Tangles of more than one strand cannot have spin 1, except if they represent composed particles. Classifying one-stranded tangles thus allows classifying all elementary gauge bosons.

Page 169

Mathematicians have already classified one-stranded tangles; they are usually called *open* or *long knots*. To get an overview, we list an example for each class of one-stranded tangles on the left-hand side of Figure 74. We will explore them in the following.

UNKNOTTED CURVES

The simplest type of tangle made of one strand is an *unknotted curve*, shown as example 1a in Figure 74. The study of gauge interactions has shown that unknotted strands are, depending on their precise shape, either vacuum strands or gauge bosons.

Page 229

In the strand model, vacuum strands are, on average, *straight*. In this property, vacuum strands differ from gauge bosons, which, on average, have *curved* strands, and thus carry energy.

GAUGE BOSONS

Gauge bosons are the carrier particles of the interactions. In the strand model, the gauge interactions are due to the three Reidemeister moves. The electromagnetic, the weak and

TABLE 12 The full list of *basic* intrinsic properties of quantum particles, from which all other observed intrinsic properties of particles, objects and images can be deduced.

PROPERTY	POSSIBLE VALUE	DETERMINES
Quantum numbers due to space-time symmetries:		
Spin S or J	integer or half-integer multiple of $\hbar$	statistics, rotation behaviour, conservation
P parity	even $(+1)$ or odd (-1)	behaviour under reflection, conservation
C parity	even $(+1)$ or odd (-1)	behaviour under charge conjugation, conservation
Interaction properties:		
Mass M	between 0 and the Planck mass	gravitation, inertia
Electric charge Q	integer multiples of one third of electron or proton charge	Lorentz force, coupling to photons, conservation
Weak charge	rational multiple of weak coupling constant	weak scattering and decays, coupling to W and Z, partial conservation
Mixing angles	between 0 and $\pi/2$	mixing of quarks and neutrinos, flavour change
CP-violating phases	between 0 and $\pi/2$	degree of CP violation in quarks and neutrinos
Strong charge, i.e., colour	rational multiple of strong coupling constant	confinement, coupling to gluons, conservation
Flavour quantum numbers, describing elementary particle content:		
Lepton number(s) L'	integer(s)	conservation in strong and e.m. interactions
Baryon number B	integer times 1/3	conservation in all interactions
Isospin I_z or I_3	$+1/2$ or $-1/2$	up and down quark content, conservation in strong and e.m. interactions
Strangeness S'	integer	strange quark content, conservation in strong and e.m. interactions
Charmness C'	integer	charm quark content, conservation in strong and e.m. interactions
Bottomness B'	integer	bottom quark content, conservation in strong and e.m. interactions
Topness T'	integer	top quark content, conservation in strong and e.m. interactions

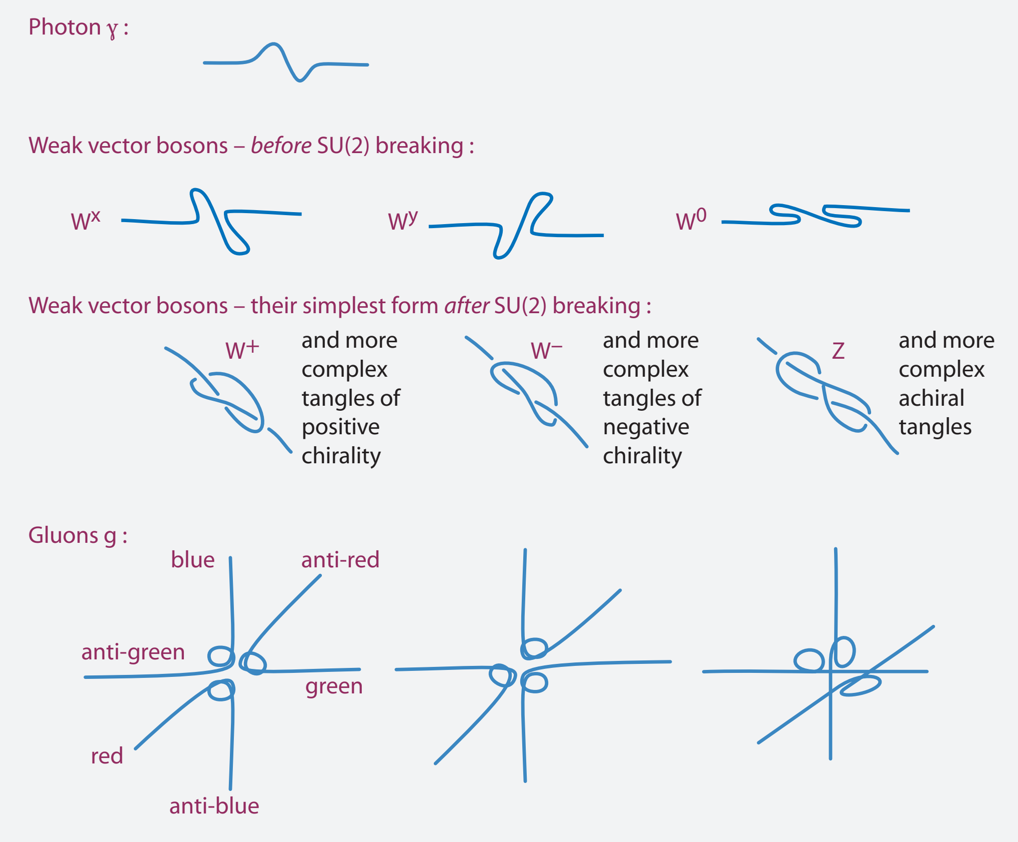


FIGURE 75 The gauge bosons in the strand model. All are made of one strand. Of the nine gluons, only eight are linearly independent.

the strong interaction correspond to respectively the first, second and third Reidemeister move. Aswe have seen above, when the three Reidemeister moves act on fermion tangles cores they generate U(1), SU(2) and SU(3) gauge symmetries. The detailed exploration of the correspondence between tangle deformation and gauge theory leads to the gauge boson tangles shown in Figure 75. Page 207

Page 211 As explained above, the *first* Reidemeister move, the twist, leads to the modelling of photons as helical strands. Therefore, photons have vanishing mass and two possible polarizations. A single unknotted strand also implies that photons do not interact among themselves: they generate an Abelian gauge theory. Automatically, photons have no weak and no strong charge. The strand model further implies that photons have negative P-parity and C-parity, as is observed.

Page 235 The study of the *third* Reidemeister move, the slide, led us to the existence of eight gluons. The eight gluons are unknotted, thus they carry no electric charge, no weak charge and no mass. Each gluon tangle has two possible polarizations. The strand model of gluons also implies that they have negative P-parity and no C-parity, as is observed. Gluons tangles carry colour and interact among themselves, thus they generate a non-Abelian gauge theory. In contrast to the other two interactions, free, single gluons are short-lived, because their structure induces rapid hadronization: when gluons act on the vacuum, quark–antiquark pairs are produced.

The next simple type of tangle made of one strand is a simple open or closed knot, or, as mathematicians say, a long or closed *prime knot*. The tangles 1b and 1b' in Figure 74 are examples. These tangles are the simplest form of the W and Z bosons after SU(2) symmetry breaking.

Page 224 The study of the *second* Reidemeister move, the *poke*, showed that deformations induced by pokes can also involve the border of space; this leads to the symmetry breaking of the weak interaction. As a result, the W and the Z boson strands are best described by knotted strands. Therefore, they W and the Z boson are massive. We have seen above Page 229 that a strand with an overhand knot is a W boson, and a strand with a figure-eight knot is a Z boson. The tangle of the W is chiral, and is thus electrically charged. The tangle of the Z is achiral and thus electrically neutral. Being knotted, the W and the Z also carry weak charge and thus interact among themselves, generating a non-Abelian, or Yang-Mills gauge theory. The strand model also implies that the W and the Z have no P-parity, no C-parity and no colour charge, as is observed.

In somewhat sloppy language we can say that the shape of photons is one-dimensional, that of the unbroken weak bosons is two-dimensional, and that of the gluons is three-dimensional. This is the essential reason that they reproduce the U(1), SU(2) and SU(3) groups, and that no higher gauge groups exist in nature.

For completeness we mention that by assignment, all gauge bosons, being made of a single strand, have vanishing lepton and baryon number, and thus also lack all flavour quantum numbers, as is observed.

COMPLICATED KNOTS

A strand can for also a highly complex knot, with a large number of crossings. We have Page 231 explained earlier on that all such possibilities – mathematically speaking, all *open knots* – are higher-order versions of the propagating W and Z bosons. They are thus due to the weak interaction. The reason for this assignment is the ability of the weak interaction to Page 229 change particle topology through overcrossing at the border of space. As a consequence, a simple particle knot can temporarily be changed into a more complex particle knot through strand fluctuations. Such complex, higher-order states include prime tangles with crossing numbers much larger than examples 1b and 1b' in Figure 74, and composed tangles, such as 1c and 1c'. In fact, *all* complex tangles of strands are due to the weak interaction. We will use this connection regularly.

In short, the strand model assigns *infinitely many* long knots to W and Z bosons. In particular, the strand model classifies all non-trivial long knots into three classes: all *achiral* long knots are assigned to the Z boson, all *positively chiral* long knots are assigned to the W^+ boson, and all *negatively chiral* long knots are assigned to the W^- boson.

CLOSED KNOTS

Figure 74 shows, on the right hand side, examples for all classes of *closed* knots, i.e., tangles *without tails*. Such objects can appear in the strand model only as *virtual* states. We thus can classify them together with their open counterparts.

We have thus two types of virtual particles. The first type are cloded knots. The second

Page 207 type are the virtual particles that we encountered in the chapter on gauge theory, where virtual particles were deformation of vacuum strands.

Summary on tangles made of one strand

In summary, all tangles made of *one strand* are *elementary* particles of spin 1, thus elementary vector bosons. Conversely, all elementary spin-1 particles are made of one strand, because other tangles do not reproduce the spin-1 behaviour under rotations: only one-stranded tangles return to the original strand after a core rotation by 2π.

In the strand model, *all* tangles of one strand are assigned to the *known* gauge bosons. Furthermore, the strand model reproduces (or predicts) the quantum numbers for each gauge boson. In short, there is *no* room for additional elementary gauge bosons.

In other words, the strand model predicts that all gauge bosons and thus all interactions are already known. We have thus a second argument – after the non-existence of other gauge groups – stating that no other gauge interaction exists in nature. (Both arguments against the existence of other gauge interactions are related, however; both are due to the three-dimensionality of space.) In particular, we find again that grand unification and supersymmetry are not allowed in nature. Page 245

Particles made of two strands

In the strand model, particle tangles can also be made of *two* strands. Examples for all classes of two-stranded tangles are listed in Figure 76. Each class has a physical particle assignment.

- The simplest tangle made of two strands is the *trivial tangle*, shown as example 2a in Figure 76. In the strand model, the trivial tangle, like all *separable* tangles, is a *composite* system. Each of the two strands can represent either the vacuum, a photon or a gluon. Simply stated, the trivial tangle is not an elementary particle.
- The simplest non-trivial tangle made of two strands is the *crossing*, shown as 2b in Figure 76. It is also separable; in the strand model, it is interpreted in the same way as the trivial tangle, because the two tangles look the same for certain observers. (It can also represent the simplest state of a d quark, as we will see below.)
- A new class of tangles are *rational tangles*, represented by example 2c. A rational tangle is a tangle that can be untangled by moving its tails *around*; they are distinct from prime and locally knotted tangles, shown as examples 2d and 2e, which require pulling the tail *through* the tangle to untangle it. Rational tangles are thus *weakly* tangled. As we will see, rational tangles represent the *graviton* and the *quarks*; we will discuss them in detail in the next two sections. More complex rational tangles are higher-order propagating states of the simpler ones.
- Another class of tangles are *prime tangles*, for which 2d is an example. Using the argument for complex one-stranded tangles, we conclude that prime tangles are higher-order propagating states, involving the weak interaction, of quarks.
- Still another class of tangles are *locally knotted tangles*, shown as example 2e. Also this class is due to higher-order propagating states of quarks that involve the weak interaction.
- Finally, *closed tangles* and *links*, such as the lower row of examples in Figure 76, and

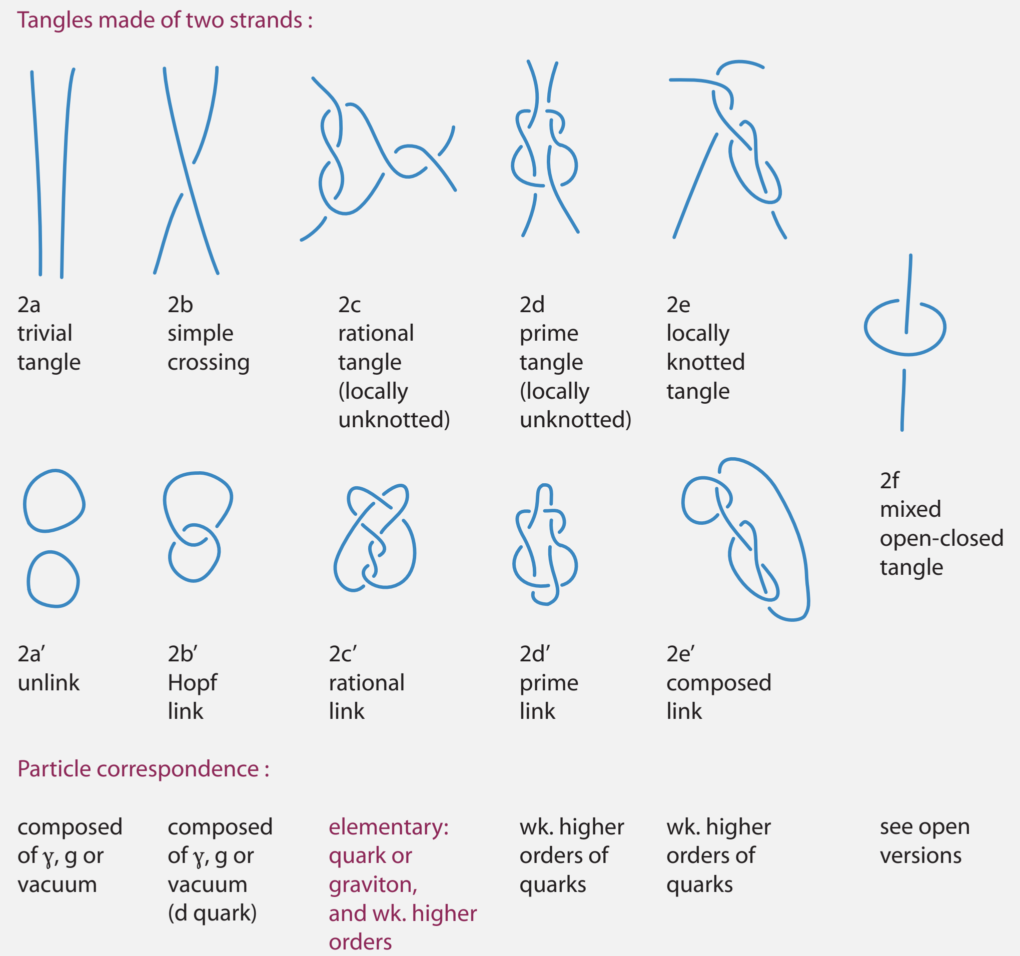


FIGURE 76 Possible tangles made of two strands.

mixed tangles, such as example 2f, are again virtual particle states.

We now explore these tangle classes in more detail.

QUARKS

Page 241 The exploration of the strong interaction earlier on has shown that the tangle of a coloured fermion, thus of a quark, must be rational, must reproduce the three possible colour options, and must break the three-belt symmetry.

The simplest tangles that realize these requirements are shown in Figure 77. The quark tangles *rational tangles* are made of *two* strands. Higher quark generations have larger crossing numbers. Two strands imply spin 1/2. The electric charges of the quarks are 1/3 and −2/3, an assignment that is especially obvious for up and down quarks and that will become even more clearer in the study of hadrons. Parity is naturally assigned as given in Figure 77. Baryon number and the other flavour quantum numbers – isospin,

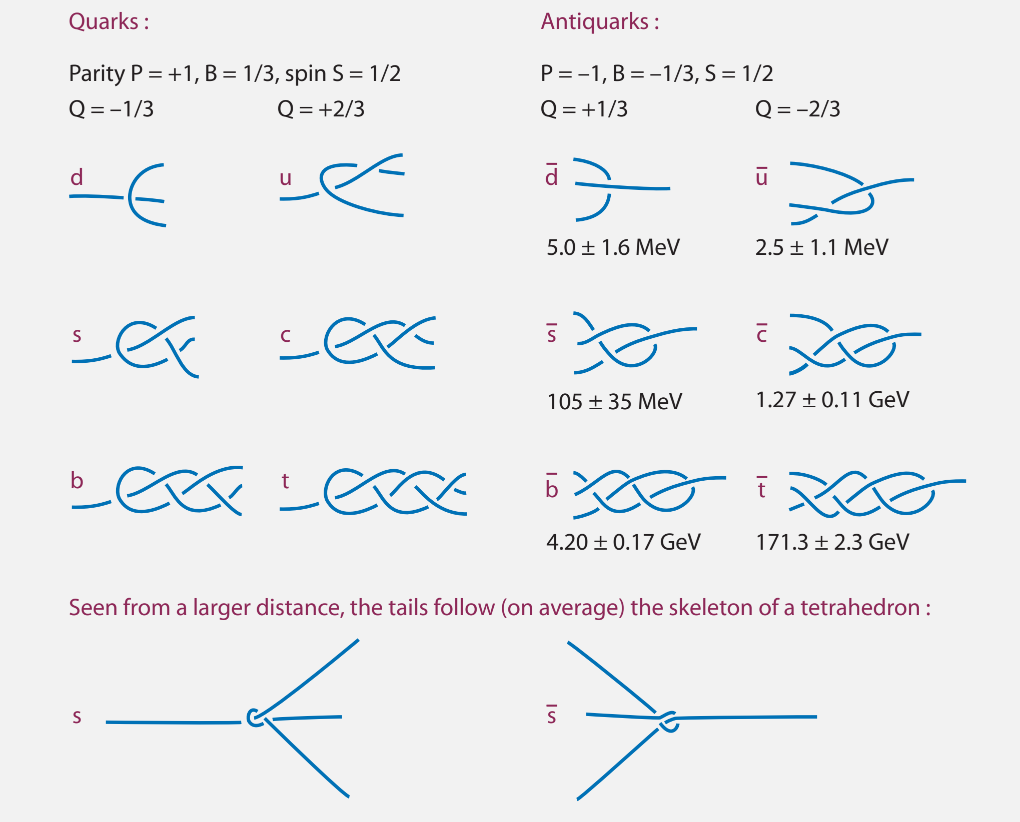


FIGURE 77 The simplest tangles assigned to the quarks and antiquarks.

strangeness, charm, bottomness, topness – are naturally assigned as usual. Flavour quantum numbers thus simply 'count' the number of corresponding quark tangles. Like all localized tangles, quarks have weak charge. We will explore weak charge in more detail below. Antiquarks are mirror tangles and have opposite quantum numbers. We will see below that these assignments reproduce the observed quantum numbers of mesons and of baryons, as well as all their other properties.

Page 318
Page 290, page 296

We note that the simplest version of the d quark is a simple crossing; nevertheless, it differs from its antiparticle, because the simple crossing mixes with the braid with seven crossings, 13 crossings, etc.; this mixing is due to the leather trick, as shown below. For all quarks, these more complex braids differ from those of their antiparticles.

Page 282

For each quark, the four tails form the skeleton of a tertrahedron. In Figure 77, the tertrahedral skeletons are drawn with one tail in the paper plane; of the other three tails, the middle one is assumed to be above the paper plane, and the outer two tails to be below the paper plane. This is important for the drawing of quark compounds. The three tails allow to reproduce the strong interaction and the colour charge of the quarks: each colour is one of the three possible orientations in space. More precisely, the three colours result from the three possible ways to map a quark tangle to the three belt structure. The quark interaction will become clearer in the section on mesons.

Page 238
Page 290

In the strand model, the quark tangles thus carry *colour*. In nature, no free coloured

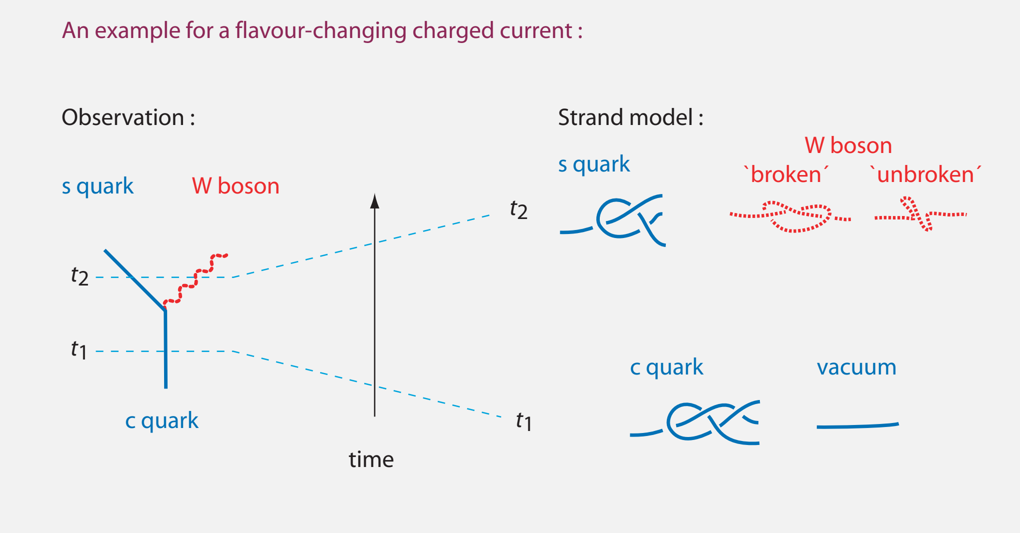


FIGURE 78 Absorption or emission of a W boson changes quark flavour.

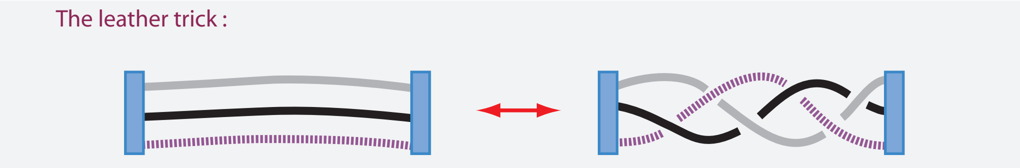


FIGURE 79 The leather trick is the deformation process that changes these two structures into each other; it limits structures made of three strands to 6 basic types.

particle has been observed. The strand model reproduces this observation in several ways. First of all, all leptons and baryons are colour-free. Secondly, only free quark tangles as shown in Figure 77 have a definite colour state, because they have a fixed orientation in space. Thirdly, a free, coloured quark tangle attracts other colour states extremely strongly, because quark tails play an important role and cannot be neglected in the strong interaction. Free quark states thus have an extremely high energy. In short, also in the strand model, only colourless composites of quarks exist as stable free particles. We will explore quark composites and the issue of confinement of quarks in more detail shortly. Page 290

Being weakly charged, quarks interact with W bosons. Indeed, the operation that takes a quark tangle and adds or subtracts a braiding step corresponds to the absorption or the emission of a W boson. This is illustrated in Figure 78. It is straightforward to check that this operation fulfils all conservation laws and properties that are observed for these so-called *flavour-changing charged currents*.

For completeness, we mention that quarks, being tangles of *two* strands, have vanishing lepton number. As we will see below, lepton tangles are made of *three* strands.

In summary, all quantum numbers of quarks are reproduced by the strand model.

FIGURE 80 The graviton in the strand model.

QUARK GENERATIONS

We stress that the quark tangles shown Figure 77 represent only the *simplest* tangle for each quark. In fact, several infinite classes of more complicated tangles are mapped to each of the six quarks. The first class is due to the *leather trick* shown in Figure 79. This trick is known to all people in the leather trade: if a braid of three strands has $n \geqslant 6$ crossings, it can be deformed into a braid with $n - 6$ crossings. Due to the leather trick, there is thus no way to introduce more than 6 quarks in the strand model. In other words, the strand model for the quarks implies that there are only 6 quarks, i.e., only three generations.

In fact, the leather trick argument assumes that the braid end can be moved *through* the braids. In the strand model, this must happen at the horizon, a region where space (and time) are not well-defined, and where such manipulations become possible. The low probability of this process will be important in the determination of quark masses. Page 308

In short, in the strand model, each quark is thus not only represented by the tangles shown in Figure 77, but also by tangles with 6 additional crossings, with 12 additional crossings, etc.

In addition, two other infinite classes of tangles are mapped to the quarks; as mentioned above, the prime and the locally knotted tangles correspond to higher-order propagators due to the weak interaction. These two infinite classes of tangles are also mapped to the six quarks. In summary, the tangle model thus leads us again to map an infinite number of tangles to the *same* particle.

As a mathematical check, we can also ask whether *all* rational tangles are mapped to quarks. It turns out that this is indeed the case. All the rational tangles that do not appear in Figure 77 are higher-order propagators of the tangles shown. Challenge 149 e

THE GRAVITON

Page 261 One rational tangle made of two strands is special. This special tangle is shown (again) in Figure 80. It differs from a quark tangle in one property: the tails are parallel (and near) to each other. Its tangle returns to its original state after rotation by π, and therefore models a spin-2 particle. The tangle is not knotted, thus has no mass, no electric, no weak and no colour charge, as expected from the graviton. Similar tangles with higher winding numbers represent higher orders in the perturbation theory of gravitation.

Page 248 The chapter on gravitation has shown how gravitons lead to curvature, horizons and the field equations of general relativity.

GLUEBALLS

Ref. 205, Ref. 206 There is no observational evidence for glueballs yet, even though lattice simulations predict the existence of several such states in the 1.5 GeV/c^2 mass range. The lack of data is usually explained by the strong background noise in the reaction that produces glueballs, and by the expected strong mixing with mesons of similar quantum numbers. The experimental search for glueballs is still ongoing.

The lowest-mass glueball is usually expected to be made of two gluons. In the strand model, a glueball made of two gluons would be made of two strands. However, the strand model of gluons does not seem to allow such a tangle. In other words, the strand model points to a SU(3) gauge theory without any bound states of the field bosons, and thus without the possibility of classical gluon waves which would correspond to electromagnetic waves.

Still, the assumed lack of glueballs in the strand model needs a more precise investigation. Ref. 207 Whatever the situation for glueballs might be, the strand model of gluons seems in contrast with the models of glueballs as knots that were proposed by Buniy and Kephart or by Niemi. Ref. 209 These models are based on closed knots, not on tangles with tails. The strand model does not seem to allow real particles, of zero spin, that are composed of gluons. However, if such tangles were somehow possible, they would imply the existence of glueballs.

THE MASS GAP PROBLEM AND THE CLAY MATHEMATICS INSTITUTE

Ref. 208 The Clay Mathematics Institute offers a large prize to anybody who proves the following statement: *For any compact simple non-Abelian gauge group, quantum gauge theory exists in continuous, four-dimensional space-time and produces a mass gap.* This is one of their so-called *millennium problems*.

The strand model does not allow arbitrary gauge groups in quantum gauge theory. According to the strand model, the only compact simple non-Abelian gauge group of interest is SU(3), the gauge group of the strong nuclear interaction. And since the strand model does not seem to allow for glueballs, for SU(3) it predicts an effective mass gap of the order of the Planck mass. (If glueballs would exist in the strand model, the mass gap would be smaller.) Indeed, the strand model explains the short range of the strong interaction as a consequence of the details of Reidemeister III moves and the quark tangle topology.

The strand model further states that space-time and gauge groups are low-energy approximations, because neither points nor fields exist at a fundamental level; points and fields are approximations to strands. According to the strand model, the *quantum* properties of nature result from the extension of strands. As a consequence, the strand model denies the existence of *any quantum* gauge theory as a separate, exact theory on *continuous* space-time.

In summary, the strand model does predict a mass gap for SU(3); but the strand model also denies the existence of quantum gauge theory for any other compact simple non-Abelian gauge group. And even in the case of SU(3) it denies – like for any other gauge groups – the existence of a quantum gauge theory on continuous space-time. As deduced Page 245 above, the strand model allows only the three known gauge groups, and allows them only

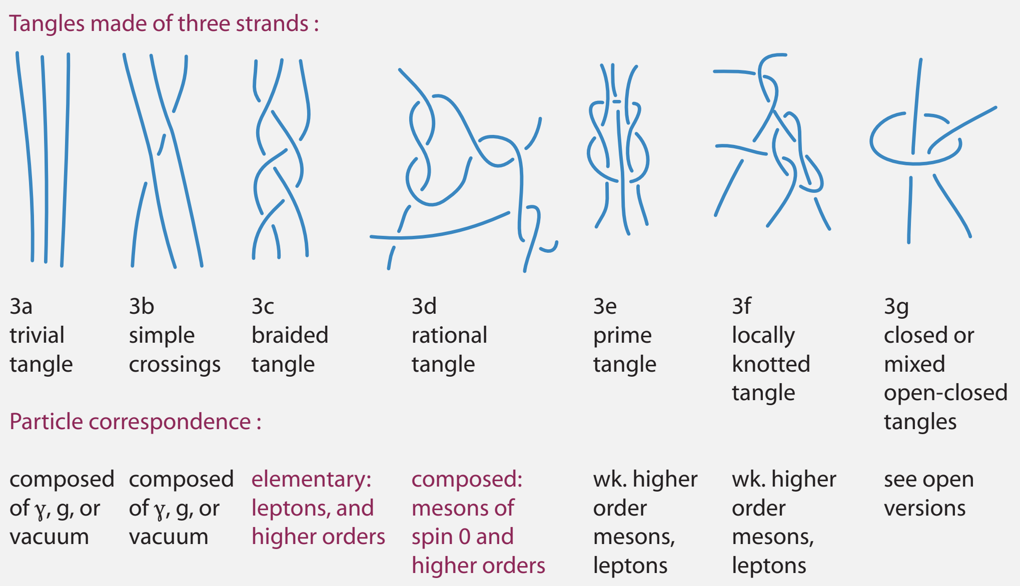


FIGURE 81 Examples for all classes of tangles made of three strands.

in the strand model of space-time.

Summary on two-stranded tangles

In summary, the strand model predicts that apart from the six quarks and the graviton, no other two-stranded elementary particles exist in nature. Concerning composite particles, the glueball issue is not completely settled, but points towards non-existence.

Quarks and the graviton, the elementary particles made of two strands, are *rational* tangles. Their strand models are thus not tangled in a complicated way, but tangled in the *least complicated* way possible. This connection will be of importance in our search for elementary particles that are still undiscovered.

Particles made of three strands

In the strand model, the next group are particles made of *three* strands. We list examples for all classes of three-stranded tangles in Figure 81. Several classes of three-stranded tangles turn out to be composites of two-stranded particles. However, a number of tangles are new and represent elementary particles.

Leptons

The candidate tangles for the leptons, given as example 3c in Figure 81, are the simplest possible non-trivial tangles with three strands. They are shown in more detail in Figure 82. The lepton tangles are simply braids with tails up to the border of space. The six tails point along the coordinate axes. These braided tangles have the following properties.

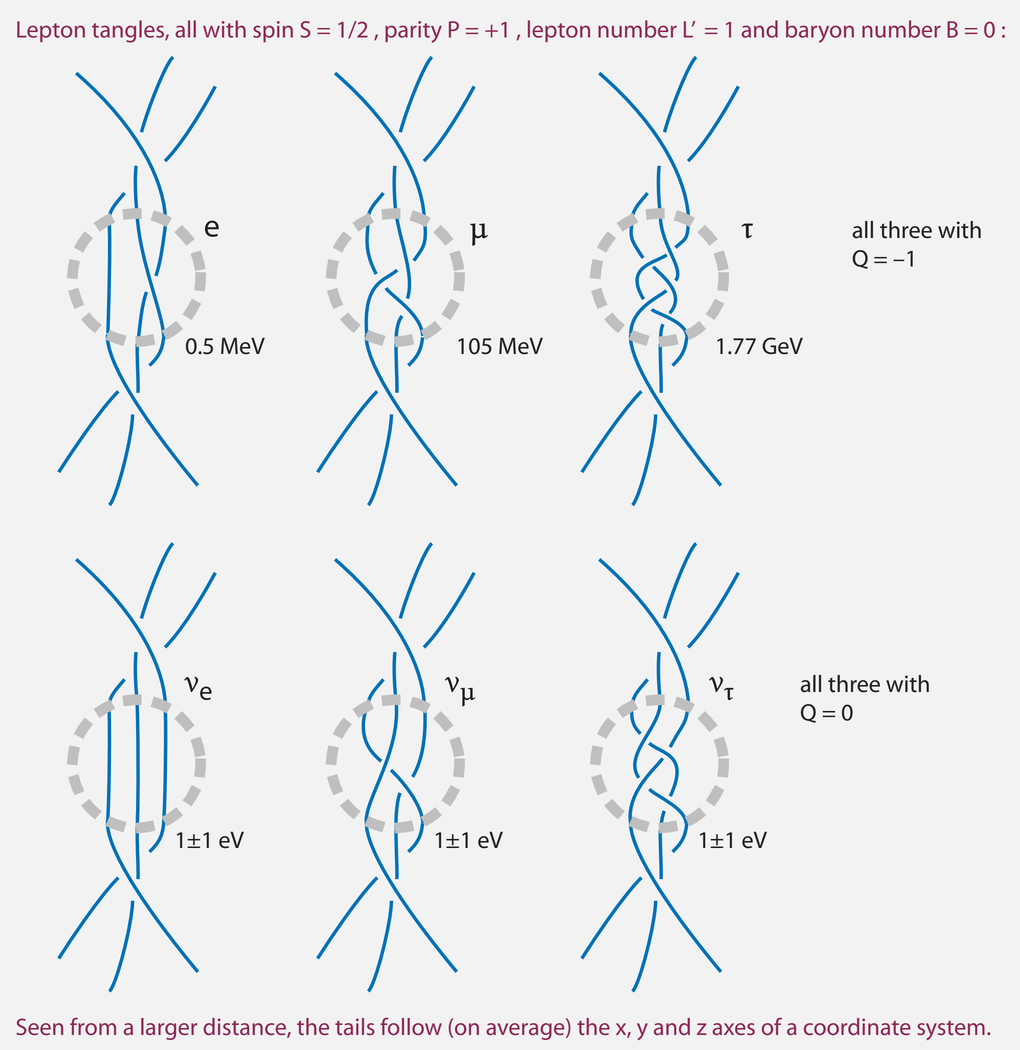


FIGURE 82 The simplest tangles of the leptons. Antileptons are mirror tangles.

- Each lepton is localized. Each lepton therefore has mass and has spin 1/2, and thus follows the Dirac equation. Each lepton has weak charge.
- Leptons and antileptons differ. In particular, neutrinos and antineutrinos differ, and both are predicted to show both chiralities.
- Three of the tangles are chiral, thus electrically charged, and three other tangles are achiral, thus uncharged.
- The spatial parity *P* of the charged lepton tangles is opposite to that of their antiparticles.
- Being made of three strands, lepton tangles have vanishing colour charge and vanishing baryon number.
- In contrast to quarks, lepton tangles are well-localized tangles and are predicted to exist as free particles.

- The three lepton (flavour) numbers can be assigned as usual; the lepton numbers are conserved in reactions, apart for neutrino mixing effects, as we will see below.
- The strand model predicts that the electron, the charged tangle with the lowest mass, is stable, as there is no way for it to decay and conserve charge and spin. The other two generations are predicted to be unstable, due to weak decays that simplify their topology.
- The number of generations is reproduced by the strand model, as every more complex braid can be seen as equivalent to one of the first six braids, with the same leather trick argument that limits the number of quarks.
- There is a natural mapping between the six quarks and the six leptons that appears if the final bend of the 'longer' quark strand is extended to the border of space, thus transforming a two-stranded quark braid into a three-stranded lepton braid. Thus we get three common generations for quarks and leptons.
- The neutrino strands differ by tail overcrossing; the strand model thus predicts that the weak interaction mixes neutrinos.
- All lepton tangles differ from each other. Thus the mass values are different for each lepton.
- Due to the small amount of tangling, the strand model predicts that the masses of the leptons are much smaller than those of the W and Z boson. This is indeed observed. (This suggests a relation between the mass and the total curvature of a tight tangle. We will explore this in detail below.)
- The simplest tangle for the electron neutrino also suggests that the mass values for the electron neutrino is naturally small, as its tangle is almost not tangled.
- The strand model also predicts that lepton masses increase with the generation number. Since the neutrino masses are not precisely known, this prediction cannot yet be checked.

In summary, tangles of three strands have precisely the quantum numbers and properties of leptons. In particular, the strand model predicts exactly three generations of leptons, and neutrinos are predicted to be Dirac particles. Ref. 210 This implies that searches for the neutrino-less double beta decay should yield negative results, that the magnetic moments of the neutrinos should have the exceedingly small values predicted by the standard model of particle physics, and that rare muon and other decays should occur at the small rates predicted by the standard model. Ref. 211

Open challenge: Find better arguments for the lepton tangles

The argument that leads to the tangles is vague. The tangles might even need small corrections. Can you improve the situation? Challenge 150 ny

The Higgs boson

The existence of the Higgs boson is predicted from the standard model of elementary particle physics using two arguments. First of all, the Higgs boson prevents unitarity violation in longitudinal W–W boson scattering. Secondly, the Higgs boson confirms the symmetry breaking mechanism of SU(2) and the related mass generation mechanism of fermions. Quantum field theory predicts that the Higgs boson has spin 0, has no electric

FIGURE 83 A candidate tangle for the Higgs boson in the strand model.

or strong charge, and has positive C and P parity. In other words, the Higgs boson is predicted to have, apart from its weak charge, the same quantum numbers as the vacuum.

In the strand model, there is only one possible candidate tangle for the Higgs boson, shown in Figure 83. The tangle has positive C and P parity, and has vanishing electric and strong charge. The tangle also corresponds to the tangle added by the leather trick; it thus could be seen to visualize how the Higgs boson gives mass to the quarks and leptons.

However, there are two issues with the candidate tangle of Figure 83. First, the tangle is a deformed, higher-order version of the electron neutrino tangle. Secondly, the spin value is not 0. In fact, there is no way at all to construct a spin-0 tangle in the strand model. These issues lead us to reconsider the arguments for the existence of the Higgs boson altogether.

Page 229 We have seen that the strand model proposes a clear mechanism for mass generation:

▷ ***Mass*** is due to strand overcrossing at the border of space.

This mechanism, due to the weak interaction, explains the W and Z boson masses. The leather trick that explains fermion masses can be seen as a sixfold overcrossing. In particular, the rarity of the overcrossing process explains why particle masses are so much smaller than the Planck mass. In short, the strand model explains mass *without* a Higgs boson.

If the Higgs does not exist, how is the unitarity of longitudinal W and Z boson scattering maintained? The strand model states that interactions of tangles in particle collisions
are described by deformations of tangles. Tangle deformations in turn are described by
Page 193 unitary operators. Therefore, the strand model predicts that unitarity is never violated
in nature. In particular, the strand model automatically predicts that the scattering of longitudinal W or Z bosons does *not* violate unitarity.

In other terms, the strand model predicts that the conventional argument about unitarity violation, which requires a Higgs boson, must be wrong. How can this be? There
Ref. 212 are at least two loopholes available in the literature, and the strand model realizes them
both.

The first known loophole is the appearance of non-perturbative effects. It is known for a long time that non-perturbative effects can mimic the existence of a Higgs boson in
usual, perturbative approximations. In this case, the standard model could remain valid
Ref. 213 at high energy without the Higgs sector. This type of electroweak symmetry breaking
would lead to longitudinal W and Z scattering that does not violate unitarity.

The other loophole in the unitarity argument appears when we explore the details of the longitudinal scattering process. In the strand model, longitudinal and transverse W

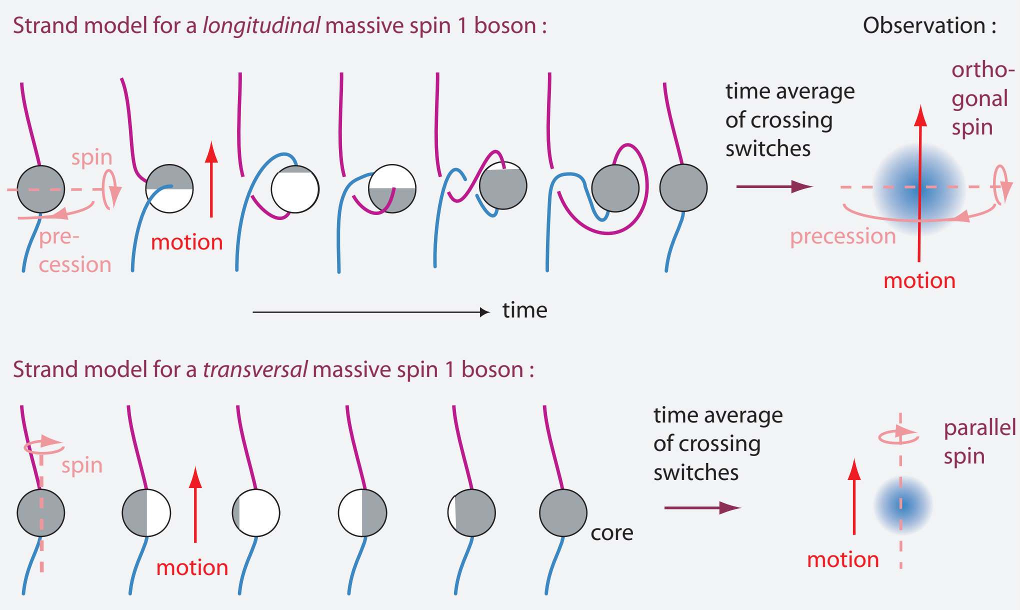


FIGURE 84 In the strand model, transverse and longitudinal W and Z bosons differ.

or Z bosons are modelled as shown in Figure 84. For longitudinal bosons, spin and its precession leads to a different situation than transversal bosons: longitudinal bosons are *more delocalized* than transversal bosons. This is not the case for fermions, where the belt trick leads to the *same* delocalization for longitudinal and transverse polarization. Interestingly, it is also known for a long time that different delocalization for longitudinal and transversal bosons *maintains* scattering unitarity, and that in the case of delocalization Ref. 214 the conventional argument for the necessity of the Higgs boson is wrong. These are well-known consequences of the so-called *non-local regularization* in quantum field theory. The strand model thus provides a specific model for this non-locality, and at the same time explains why it *only* appears for longitudinal W and Z bosons.

In other words, the strand model predicts that the scattering of longitudinal W and Z bosons is the first system that will show effects specific to the strand model. Such precision scattering experiments will be possible soon at the Large Hadron Collider in Geneva. These experiments will allow to check the *non-perturbative effects* and the *non-local effects* predicted by the strand model. For example, the strand model predicts that the wave function of a longitudinal and a transversally polarized W or Z boson of the same energy differ; this should yield different cross sections.

In summary, the strand model predicts well-behaved scattering amplitudes for longitudinal W and Z boson scattering in the TeV region, together with the absence of the Higgs boson.[*] The strand model explains mass generation and lack of unitarity viola-

---

[*] If the arguments against the existence of the Higgs boson turn out to be wrong, then the strand model might be saved with a dirty trick, namely by arguing that the tangle of Figure 83 might effectively have spin 0. This dirty trick would predict *exactly one* Higgs boson. In this case, the ropelength of the Borromean rings, 29.03, would lead to a Higgs mass prediction of 114 GeV. If, however, two or more Higgs bosons were found to exist in nature, the strand model would be definitely falsified.

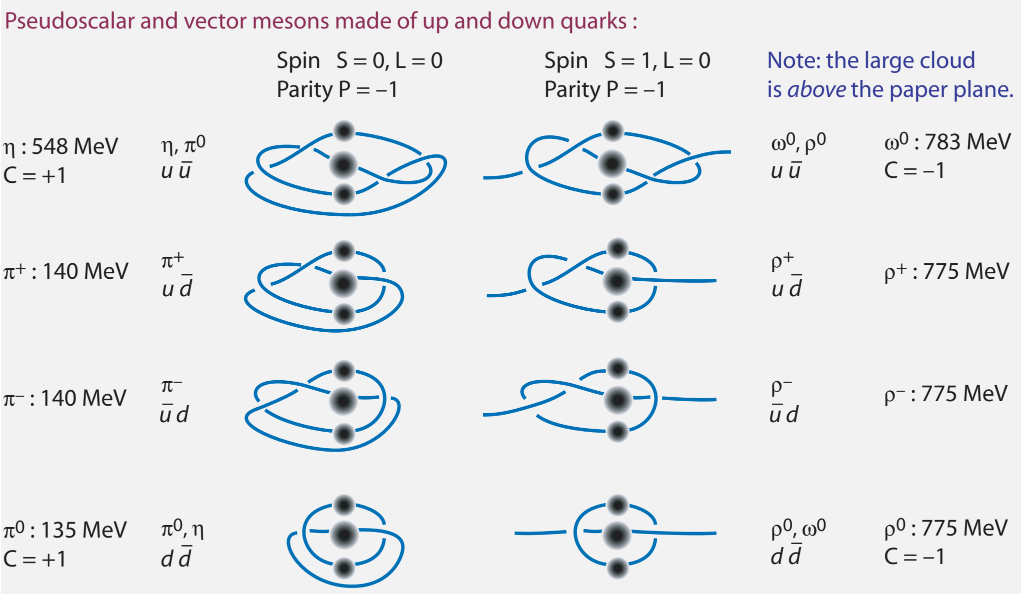


FIGURE 85 The simplest strand models for the light pseudoscalar and vector mesons (clouds indicate crossed tail pairs to the border of space).

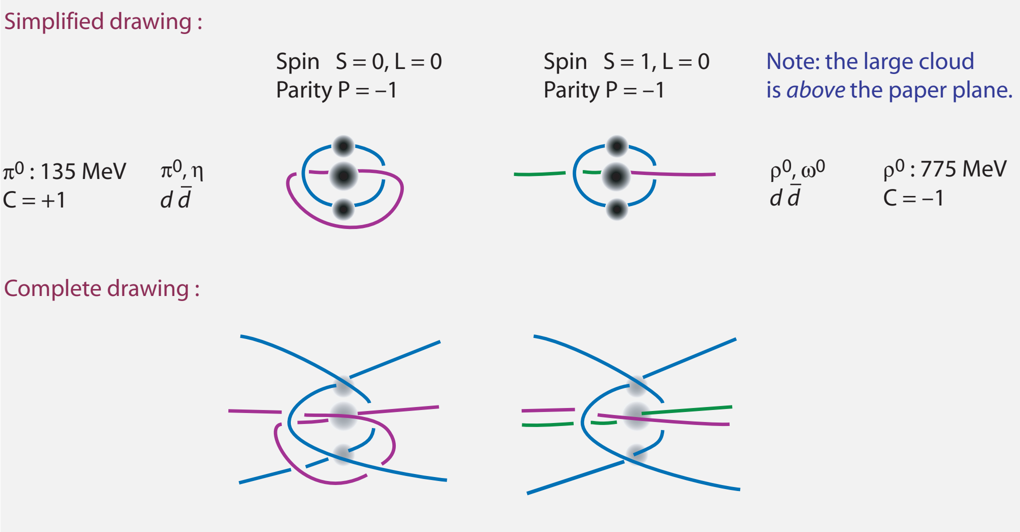


FIGURE 86 The meaning of the clouds used in the tangle graphs of mesons and baryons.

tions in longitudinal W or Z boson scattering as consequences of overcrossing, i.e., as non-perturbative and non-local effects, and not as consequences of an elementary spin-0 Higgs boson. The forthcoming experiments at the Large Hadron Collider in Geneva will test this prediction.

Quark-antiquark mesons

In the strand model, all three-stranded tangles apart from the leptons represent *composite* particles. The first example are *mesons*.

In the strand model, rational tangles of three strands are quark-antiquark mesons with spin 0. The quark tangles yield a simple model of these *pseudoscalar* mesons, shown on the left-hand sides of Figure 85, Figure 87 and Figure 88. The right-hand sides of the figures show vector mesons, thus with spin 1, that consist of *four* strands. All tangles are rational. Inside mesons, quarks and antiquarks 'bond' at three spots that form a triangle oriented perpendicularly to the bond direction and to the paper plane. To increase clarity, the 'bonds' are drawn as clouds in the figures; however, they consist of two crossed (linked) tails of the involved strands that reach the border of space, as shown in Figure 86.

With this construction, only mesons of the form $\overline{q}\,q$ are possible. Other combinations,
Ref. 215 such as $q\,q$ or $\overline{q}\,\overline{q}$, turn out to be unlinked. We note directly that this model of mesons
Ref. 216 resembles the original string model of hadrons from 1973, the Lund string model and
Ref. 217 the recent QCD string model.

To compare the meson structures with experimental data, we explore the resulting quantum numbers. As in quantum field theory, also in the strand model the parity of a particle is the product of the intrinsic parities and of wave function parity. The states with orbital angular momentum $L = 0$ are the lowest states. Experimentally, the lightest mesons have quantum numbers $J^{PC} = 0^{-+}$, and thus are pseudoscalars, or have $J^{PC} = 1^{--}$, and thus are vector mesons. The strand model reproduces these observed quantum numbers. (We note that the spin of any composite particle, such as a meson, is low-energy quantity; to determine it from the composite tangle, the tails producing the bonds – drawn as clouds in the figures – must be neglected. As a result, the low-energy spin of mesons and of baryons is correctly reproduced by the strand model.)

In the strand model, the meson states are colour-neutral, or 'white', by construction, because the quark and the antiquark, in all orientations, always have opposite colours that add up to white.

In the strand model, the electric charge is an integer for all mesons. Chiral tangles are charged, achiral tangles uncharged. The charge values deduced from the strand model thus reproduce the observed ones.

In experiments, no mesons of quantum numbers 0^{--}, 0^{+-}, or 1^{-+} are observed. Also this observation is reproduced by the quark tangles, as is easily checked by direct inspection. The strand model thus reproduces the very argument that once was central to the acceptance of the quark model itself.

It is important to realize that in the strand model, each meson is represented by a *tangle family* consisting of *several* tangle structures. This has three reasons. First, the 'clouds' can be combined in different ways. For example, both the $u\,\overline{u}$ and the $d\,\overline{d}$ have as alternate structure a line plus a ring. This common structure is seen as the underlying reason that these two quark structures *mix*, as is indeed observed. (The same structure is also possible for $s\,\overline{s}$, and indeed, a full description of these mesons must include mixing with this state as well.) The second reason that mesons have several structures are the mentioned, more complex braid structures possible for each quark, namely with 6, 12, etc. additional braid crossings. The third reason for several tangles are the higher-order Feynman diagrams of the weak interaction, which add yet another group of more

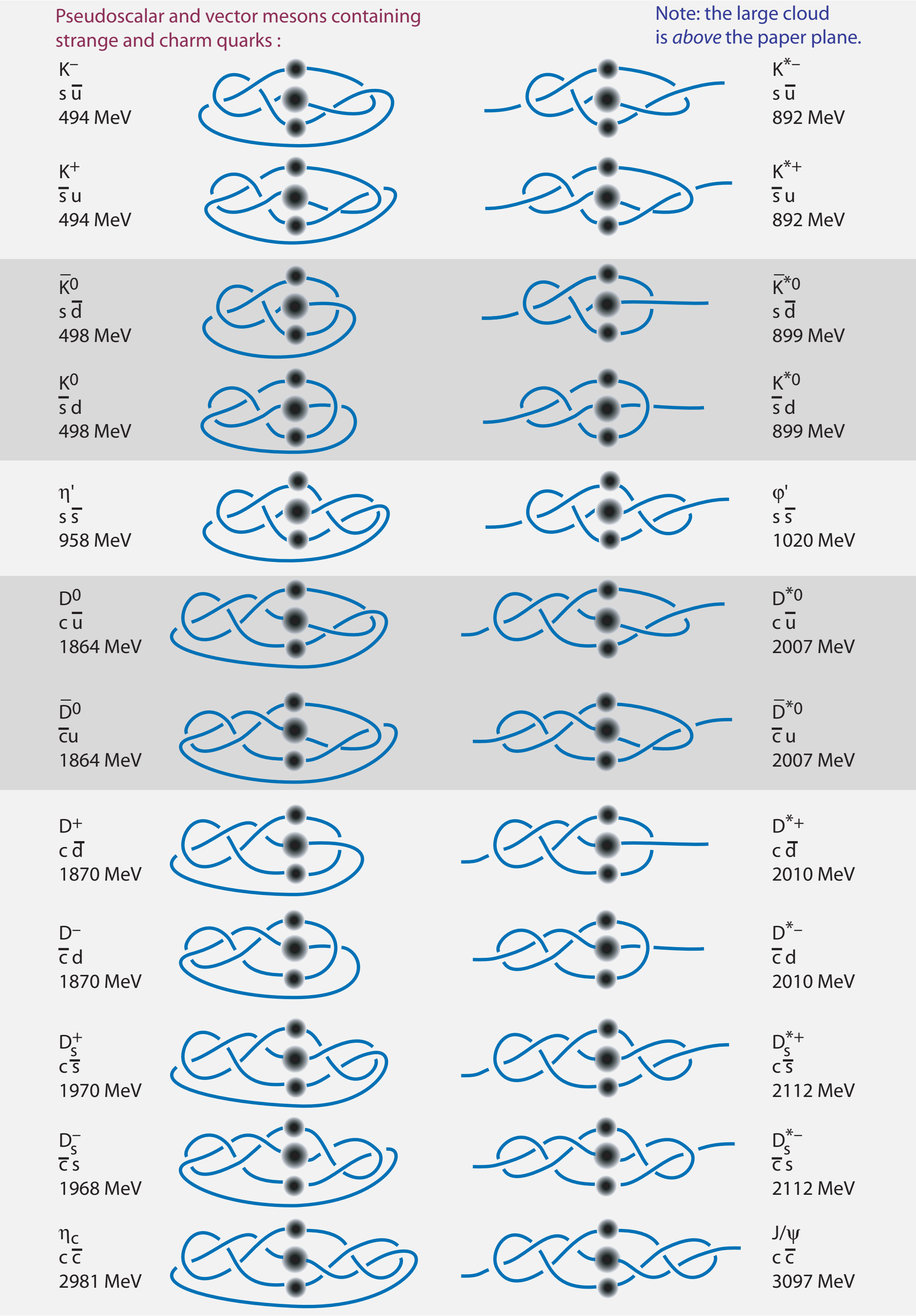


FIGURE 87 The simplest strand models for strange and charmed mesons with zero angular momentum. Mesons on the left side have spin 0 and negative parity; mesons on the right side have spin 1 and negative parity. Clouds indicate crossed tail pairs to the border of space; grey boxes indicate tangles that mix with their antiparticles and which are thus predicted to show CP violation.

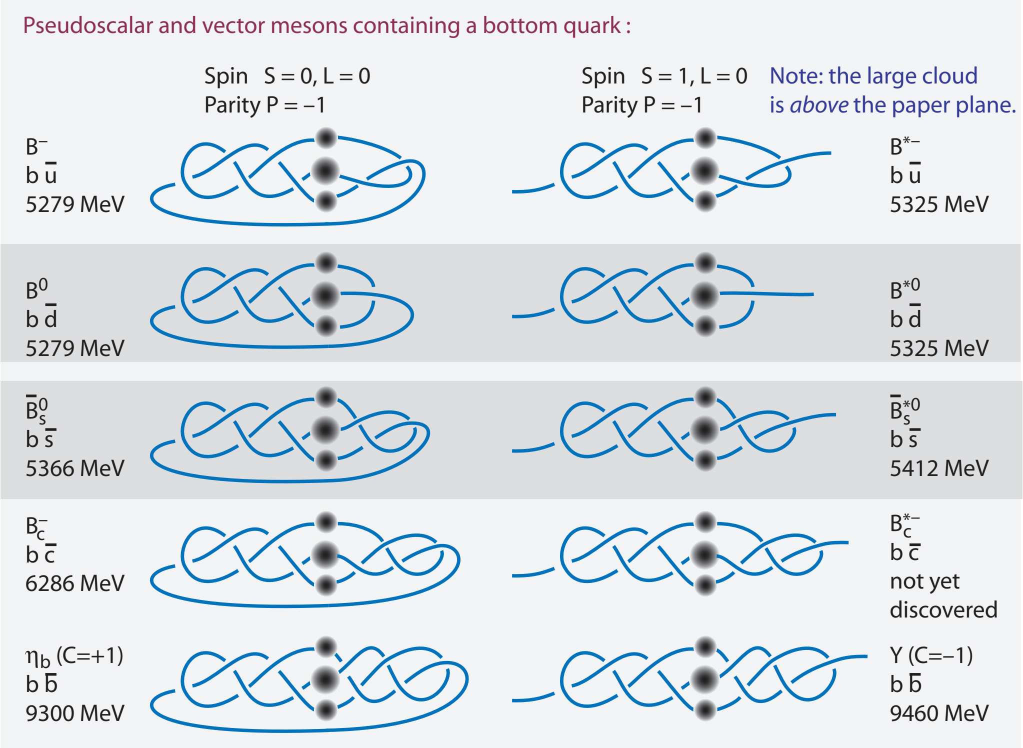


FIGURE 88 The simplest strand models for some heavy pseudoscalar and vector mesons. Antiparticles are not drawn; their tangles are mirrors of the particle tangles. Clouds indicate crossed tail pairs to the border of space; grey boxes indicate tangles that mix with their antiparticles and which are thus predicted to show CP violation.

complex topologies that also belong to each meson.

In short, the mesons structures of Figure 85, Figure 87 and Figure 88 are only the *simplest* tangles for each meson. Nevertheless, all tangles, both the simplest and the more complex meson tangles, reproduce spin values, parities, and all the other quantum numbers of mesons. Indeed, in the strand model, the more complex tangles automatically share the quantum numbers of the simplest one.

Meson form factors

The strand model also predicts directly that all mesons from Figure 85, Figure 87 and Figure 88, in fact all mesons with vanishing orbital momentum, are *prolate*. This (unsurprising) result is agreement with observations. Mesons with non-vanishing orbital momentum are also predicted to be prolate. The latter prediction is made also by all other meson models, but has not yet been checked by experiment. Ref. 219

There is another way to put what we have found so far. The strand model makes the following prediction: When the meson tangles are averaged over time, the crossing densities reproduce the measured spatial, quark flavour, spin and colour part of the meson wave functions. This prediction can be checked against measured form factors and against lattice QCD calculations.

MESON MASSES, EXCITED MESONS AND QUARK CONFINEMENT

The strand model also allows to understand meson masses. We recall that a *topologically complex* tangle implies a *large* mass. With this relation, Figure 85 predicts that the π^0, η and $\pi^{+/-}$ have different masses and follow the observed meson mass sequence $m(\pi^0) < m(\pi^{+/-}) < m(\eta)$. The other mass sequences can be checked with the help of Figure 85, Figure 87 and Figure 88; there are no contradictions with observations. However, there is one limit case: the strand model predicts different masses for the ρ^0, ω, and $\rho^{+/-}$. So far, observations only partly confirm the prediction. Recent precision experiments seem to suggest that ρ^0 and $\rho^{+/-}$ have different mass; this result has not been confirmed yet. Ref. 218

More precise mass determinations will be possible with numerical calculations. This will be explored in more detail later on. Page 306 In any case, the strand model for mesons suggests that the quark masses are not so important for the determination of meson masses, whereas the details of the quark-antiquark bond are. Indeed, the light meson and baryon masses are much higher than the masses of the constituent quarks.

The relative unimportance of quark masses for many meson masses is also confirmed for the case of *excited* mesons, i.e., for mesons with orbital angular momentum L. It is well known that mesons of non-vanishing orbital angular momentum can be grouped into sets which have the same quark content, but different total angular momentum $J = L + S$. These families are observed to follow a well-known relation between total angular momentum J and mass m, called *Regge trajectories*:

$$J = \alpha_0 + \alpha_1 m^2 \tag{202}$$

Ref. 220 with an (almost) constant factor α_1 for all mesons, about 0.9 GeV/fm. These relations, the famous *Regge trajectories*, are explained in quantum chromodynamics as deriving from the linear increase with distance of the effective potential between quarks, thus from the properties of the relativistic harmonic oscillator. The linear potential itself is usually seen as a consequence of a fluxtube-like bond between quarks.

In the strand model, the fluxtube-like bond between the quarks is built-in automatically, as shown in Figure 89. All mesons have three connecting 'bonds' and these three bonds can be seen as forming one common string tube. In the simplified drawings, the bond or string tube is the region containing the clouds. In orbitally excited mesons, the three bonds are expected to lengthen and thus to produce additional crossing changes, thus additional effective mass. The strand model also suggests a *linear* relation. Since the mechanism is expected to be similar for all mesons, which all have three bonding clouds, the strand model predicts the *same* slope for all meson (and baryon) Regge trajectories. This is indeed observed.

In summary, the strand model reproduces meson mass sequences and quark confinement in its general properties.

CP VIOLATION IN MESONS

Ref. 221 In the weak interaction, the product CP of C and P parity is usually conserved. However, rare exceptions are observed for the decay of the K^0 meson and in various processes that involve the B^0 and B^0_s mesons. In each of these exceptions, the meson is found to

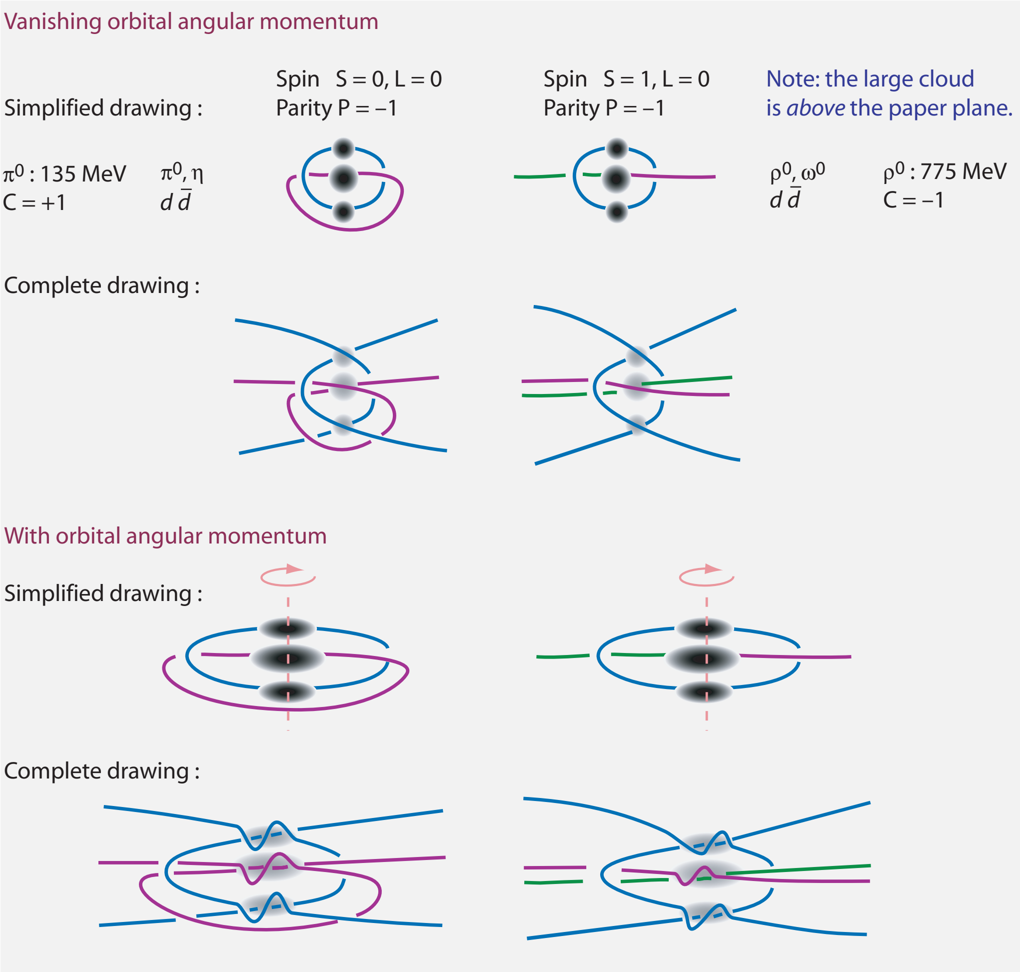


FIGURE 89 The strand model for mesons without (top) and with (bottom) orbital angular momentum.

Ref. 218 mix with its own antiparticle. CP violation is essential to explain the matter–antimatter asymmetry of the universe.

The strand model allows us to deduce whether the mixing of a meson with its own antiparticle is possible or not. As expected, only neutral mesons are candidates for such mixing, because of charge conservation. In the strand model, particle–antiparticle mixing is possible whenever the transition from a neutral meson to its antiparticle is possible in *two* ways: by taking the mirror of the meson tangle or by shifting the position of the binding strands. All mesons for which this is possible are shown in grey boxes in Figure 85, Figure 87 and Figure 88. The strand model also makes it clear that such mixing requires shifting of the bonds; this is a low-probability process that is due to the weak interaction. The strand model thus predicts that the weak interaction violates CP invariance in mesons that mix with their antiparticles.

Since the spin 1 mesons decay strongly and thus do not live long enough, the small effect of CP violation is de facto only observed in pseudoscalar, spin-0 mesons. The

Ref. 218

strand model thus predicts observable mixings and CP violation for the mesons pairs $K^0 - \overline{K}^0$, $D^0 - \overline{D}^0$, $B^0 - \overline{B}^0$, $B^0_s - \overline{B}^0_s$. The prediction by the model corresponds precisely to those systems for which CP violation is actually observed, or, as in the case of $D^0 - \overline{D}^0$, expected to be observed when the experimental difficulties will be overcome. This should happen in the coming years.

In the strand model, meson–antimeson mixing is possible because the various quarks are braided strands. Because of this braid structure, the existence of meson–antimeson mixing is a consequence of the existence of three quark generations. The meson structures also make it clear that such mixings would not be possible if there were no third quark generation. The strand model thus reproduces the usual explanation of CP violation as the result of three quark generations.

Page 304

For the strong and the electromagnetic interaction, the strand model predicts that there is no mixing and no CP violation, because gluons and photons do not change particle topology. Therefore, the strand model suggests the absence of axions. The lack of a suitable tangle for axions, shown later on, then turns this suggestions into a prediction.

In summary, the existence of CP violation in the weak interactions and the lack of CP violation in the strong interaction are natural consequences of the strand model.

OTHER THREE-STRANDED TANGLES AND GLUEBALLS

In the strand model, complicated tangles made of three strands are either higher-order propagating versions of the tangles just presented or composites of one-stranded or two-stranded particles. For example, intrinsically knotted or prime tangles made of three strands are all due to weak processes of higher order acting on elementary tangles.

Page 283

The conjectured glueballs could also be made of three gluons. In the strand model, such a structure would be a simple tangle made of three strands. However, the masslessness of gluons does not seem to allow such a tangle. The argument is not watertight, however, and the issue is still topic of research.

SUMMARY ON THREE-STRANDED TANGLES

Compared to two-stranded tangles, one *new* class of *elementary* particles appears for three strands; the new class is somewhat less tangled than general rational tangles but still more tangled than the trivial vacuum tangle: the *braided tangles*. Braided tangles represent leptons; the tangles reproduce all the observed quantum numbers of leptons. The braided tangles also imply that neutrinos and anti-neutrinos differ, are massive, and are Dirac particles.

The strand model also predicts that apart from the six leptons, no other elementary particles made of three strands exist in nature. No Higgs boson exists. (If at all, at most one Higgs boson can exist.)

In the case of *composite* particles made of three strands, the strand model proposes tangles for all pseudoscalar mesons; the resulting quantum numbers and mass sequences match the observed values. Among the composite particles, the glueball issue is not completely settled.

TANGLES OF FOUR AND MORE STRANDS

If we add one or more strand to a three-strand tangle, *no additional class* of tangles appears. The tangle classes remain the same as in the three-strand case. In other words, no additional elementary particles appear. To show this, we start our exploration with the *rational* tangles.

We saw above that the rational tangles made of four strands represent the vector mesons. We have already explored them together with the scalar mesons. But certain more complex rational tangles are also important in nature, as we consist of them.

BARYONS

In the strand model, rational tangles made of five or six strands are baryons. The quark tangles of the strand model yield the tangles for baryons in a natural way, as Figure 90 shows. Again, not all quark combinations are possible. First of all, quark tangles do not allow mixed $q\,q\,\overline{q}$ or $q\,\overline{q}\,\overline{q}$ structures, but only $q\,q\,q$ or $\overline{q}\,\overline{q}\,\overline{q}$ structures. In addition, the tangles do not allow (fully symmetric) spin 1/2 states for $u\,u\,u$ or $d\,d\,d$, but only spin 3/2 states. The model also naturally predicts that there are only two spin 1/2 baryons made of u and d quarks. All this corresponds to observation. The tangles for the simplest baryons are shown in Figure 90.

The electric charges of the baryons are reproduced. In particular, the tangle topologies imply that the proton has the same charge as the positron. Neutral baryons have topologically achiral structures; nevertheless, the neutron differs from its antiparticle, as can be deduced from Figure 90, through its three-dimensional shape. The Δ baryons have different electric charges, depending on their writhe. Page 320

Baryons are naturally colour-neutral, as observed. The model also shows that the baryon wave function usually cannot be factorized into a spin and quark part: the nucleons need *two* graphs to describe them, and tangle shapes play a role. Baryon parities are reproduced; the neutron and the antineutron differ. All this corresponds to known baryon behaviour. Also the observed oblate baryon shapes (in other words, the baryon quadrupole moments) are reproduced by the tangle model. Ref. 219

The particle masses of proton and neutron differ, because their topologies differ. However, the topological difference is 'small', as seen in Figure 90, so the mass difference is small. The topological difference between the various Δ baryons is even smaller, and indeed, their mass difference is barely discernible in experiments.

The strand model naturally yields the baryon octet and decuplet, as shown in Figure 91 and Figure 92. In general, complex baryon tangles have higher mass than simpler ones, as shown in the figures; this is also the case for the baryons, not illustrated here, that include other quarks. And like for mesons, baryon Regge trajectories are due to 'stretching' and tangling of 'cloudy' strands. Since the bonds to each quark are again (at most) three, the model qualitatively reproduces the observation that the Regge slope for all baryons is the same and is equal to that for mesons. We note that this also implies that the quark masses play only a minor role in the generation of hadron masses; this old result from QCD is thus reproduced by the strand model.

The arguments presented so far only reproduce mass sequences, not mass values. Actual hadron mass calculations are possible with the strand model: it is necessary to compute the number of crossing changes each tangle produces. There is a chance, but no

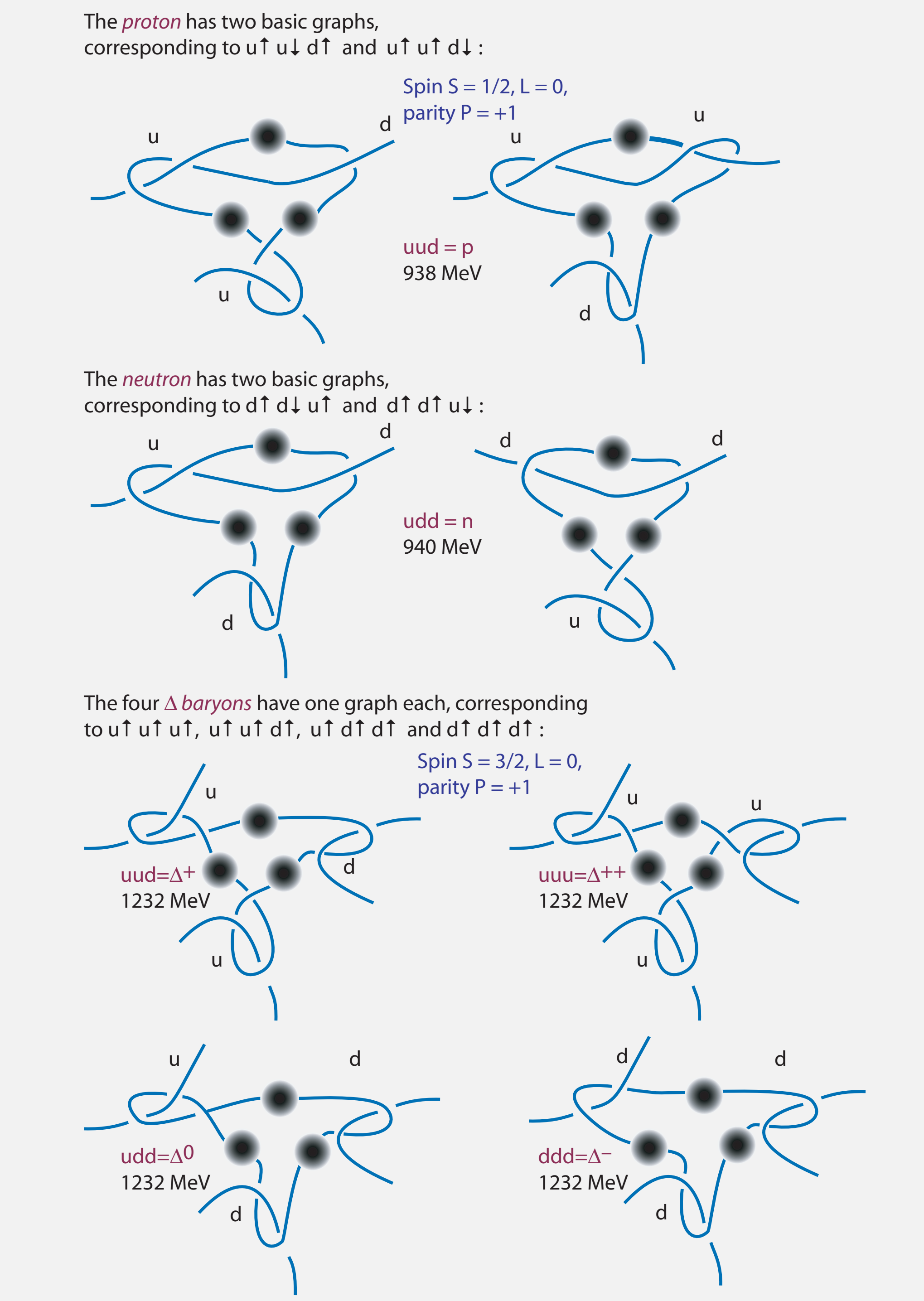


FIGURE 90 The simplest strand models for the lightest baryons made of up and down quarks (clouds indicate linked tail pairs to the border of space).

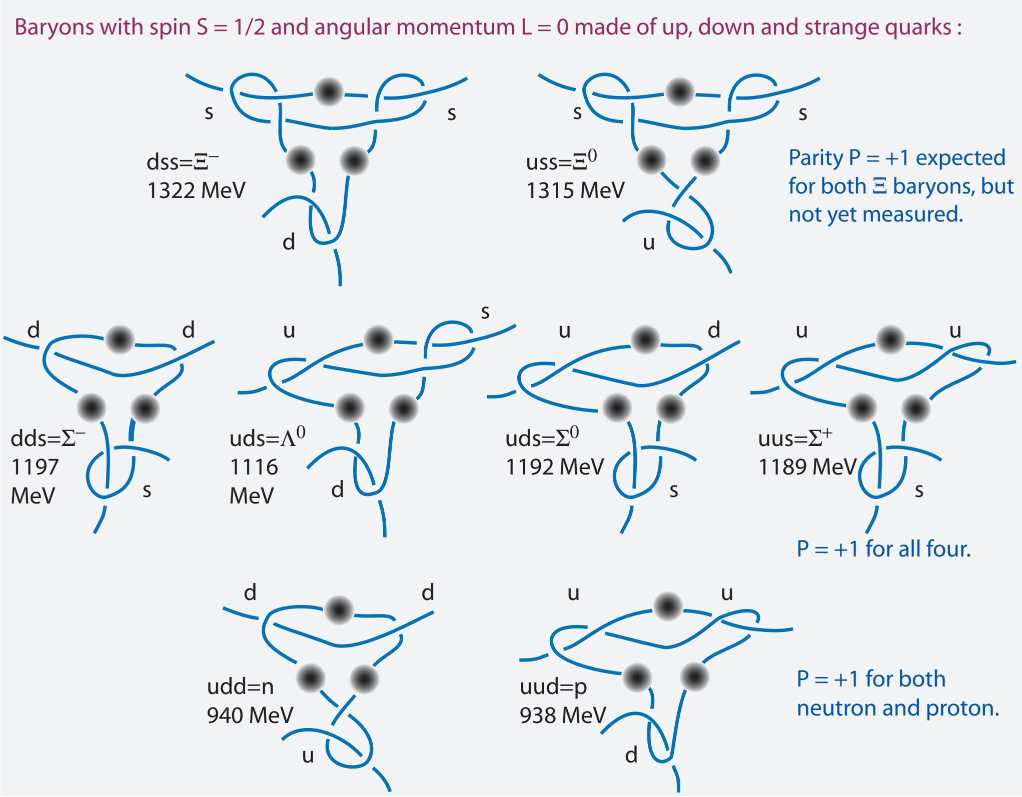


FIGURE 91 One tangle (only) for each baryon in the lowest J=L+S=1/2 baryon octet (clouds indicate linked tail pairs to the border of space).

certainty, that such calculations might be simpler to implement than those of lattice QCD.

Tetraquarks and exotic mesons

Ref. 206 Ref. 222 Among the exotic mesons, tetraquarks are the most explored cases. It is now widely believed that the low-mass scalar mesons are tetraquarks. In the strand model, tetraquarks are possible; an example is given in Figure 93. This is a six-stranded rational tangle. Spin, parities and mass sequences from the strand model seem to agree with observations. The details of this topic are left for future exploration.

Ref. 209 The strand model makes an additional statement: knotted strings in quark–antiquark states are impossible. Such states have been proposed by Niemi. In the string model, such states would not be separate mesons, but usual mesons with one or several added virtual weak vector bosons. This type of exotic mesons is therefore predicted not to exist.

Page 283, page 295 The situation for glueballs, which are another type of exotic mesons, has already been discussed above. They probably do not to exist.

Other tangles made of four or more strands

We do not need to explore other prime tangles or locally knotted tangles made of four or more strands. They are higher-order versions of rational tangles, as explained already

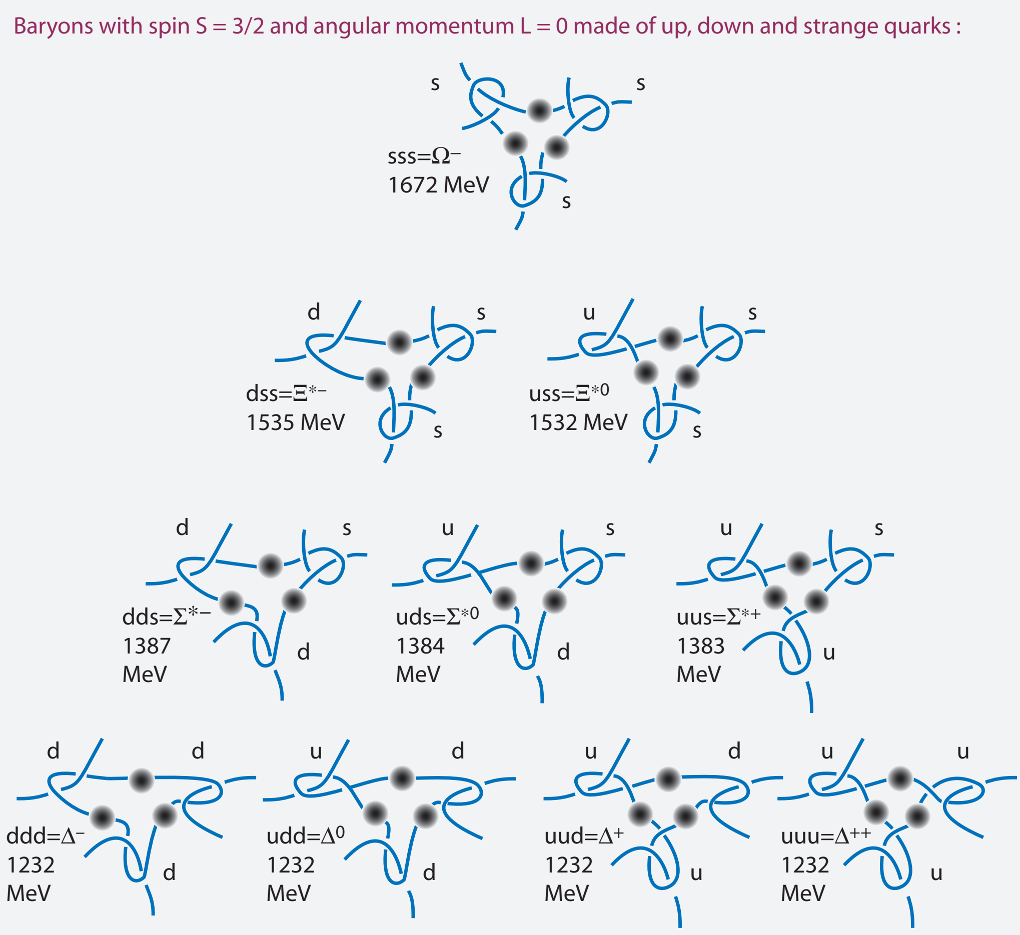


FIGURE 92 One tangle for each baryon in the lowest J=3/2 baryon decuplet (clouds indicate linked tail pairs to the border of space).

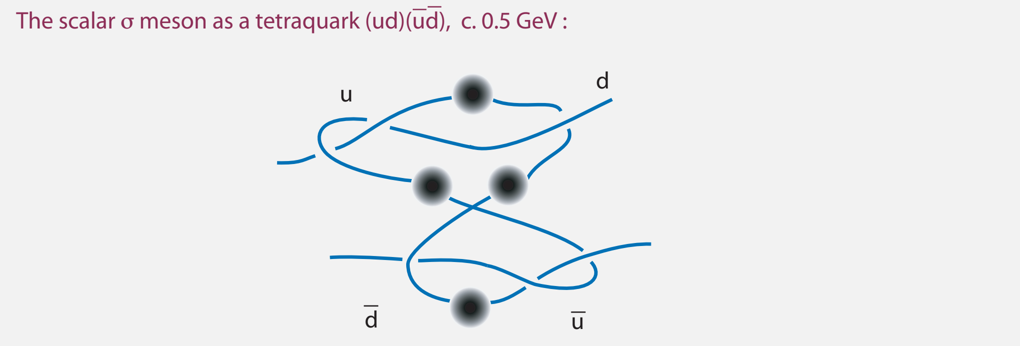


FIGURE 93 The strand model for a specific tetraquark (clouds indicate linked tail pairs to the border of space).

in the case of two and three strands. We also do not need to explore separable tangles. Separable tangles are composite of tangles with fewer strands.

One class of tangles remains to be discussed: braided tangles of four or more strands. Now, a higher-order perturbation of the weak interaction can always lead to the entanglement of some vacuum strand with a tangle of fewer strands. Braided tangles of four or more strands are thus higher-order propagating states of three-stranded leptons or hadrons.

We can also state this in another way. There are no tangles of four or more strands that are more tangled than the trivial tangle but less tangled than the lepton tangles. Therefore, no additional elementary particles are possible. In short, *the tangle model does not allow elementary particles with four or more strands.*

SUMMARY ON TANGLES MADE OF FOUR OR MORE STRANDS

By exploring all possible tangle classes in detail, we have shown that *every* localized structure made of strands has an interpretation in the strand model. In particular, the strand model makes a simple statement on any tangle made of four or more strands: such a tangle is *composite* of the elementary tangles made of one, two or three strands. In other terms, there are *no* elementary particles made of four or more strands in nature.

The strand model thus states that each possible tangle represents a physical state: an overview is given in Table 13. The mapping between tangles and particles is only possible because (infinitely) many tangles are assigned to each massive elementary particle. This, in turn, is a consequence of the topology changes induced by the weak interaction. As a result, the strand model is able to deduce that the number of elementary particles in nature is limited to those contained in the standard model of particle physics.

FUN CHALLENGES AND CURIOSITIES ABOUT PARTICLE TANGLES

In the strand model, mass appears due to overcrossing at the border of space. But mass is also due to tangle rotation and fluctuation. How do the two definitions come together?

Challenge 151 ny

* *

In the strand model, only crossing changes are observable. How then can the specific tangle structure of a particle have any observable effects? In particular, how can quantum numbers be related to tangle structure, if the only observables are due to crossing changes? Challenge 152 e

* *

Ref. 223 What is the relation of the model shown here to the ideas of Viro and Viro on skew lines?

* *

The most prominent proponent of the idea that particles might be knots was, in 1868, William Thomson–Kelvin. He proposed the idea that different atoms might be differently 'knotted vortices' in the 'ether'. The proposal was ignored – and rightly so – because it did not explain anything: neither the properties nor the interactions of atoms were explained. The proposal simply had no relation to reality. Ref. 224

* *

TABLE 13 The match between tangles and particles in the strand model.

Strands	Tangle	Particle	Type
1	unknotted	elementary	vacuum, photon, gluon
1	knotted	elementary	W or Z boson, including higher orders due to the weak interaction
2	unknotted	composed	made of simpler tangles: vacuum, photons, gluons
2	rational	elementary	quark or graviton, including higher orders
2	knotted	elementary	quarks with higher orders due to the weak interaction
3	unknotted	composed	made of simpler tangles: vacuum, photons, gluons
3	braided	elementary	leptons
3	rational	elementary or composed	leptons with higher orders due to the weak interactions, or mesons
3	knotted	elementary or composed	leptons or mesons with higher orders due to the weak interactions
4 & more	like for 3 strands	all composed	made of simpler tangles

Purely topological models for elementary particles have been proposed and explored by various scholars. But only a few researchers ever proposed specific topological structures for each elementary particle. Such proposals are easily criticized, so that it is easy to make a fool of oneself; any such proposals thus needs courage.

- Ref. 225 — Jehle modelled elementary particles as closed knots already in the 1970s. However, his model did not reproduce quantum theory, nor does it reproduce all particles known today.
- Ref. 146 — Ng has modelled mesons as knots. There is however, no model for quarks, leptons or bosons, nor a description for the gauge interactions.
- Ref. 226 — Mongan has modelled elementary particles as made of three strands that carry electric charge. However, there is no connection with quantum field theory or general relativity.
- Ref. 142 — Avrin has modelled hadrons and leptons as Moebius bands, and interactions as cut-and-glue processes. The model however, does not explain the masses of the particles or the coupling constants.
- Ref. 144 — Finkelstein has modelled fermions as knots. This approach, however, does not explain the gauge properties of the interactions, nor most properties of elementary particles.
- Ref. 143 — Bilson-Thompson and his colleagues have modelled elementary fermions and bosons as structures of triple ribbons. The leather trick is used, like in the strand model, to explain the three generations of quarks and leptons. This is by far the most complete model from this list. However, the origin of particle mass, of particle mixing and,

most of all, of the gauge interactions is not clear yet.

* *

Strands are not superstrings. The fundamental principle of the strand model is not fulfilled by superstrings. In contrast to superstrings, strands have no tension, no supersymmetry and no own Lagrangian. Since strands have no tension, they cannot oscillate. In the strand model, particles are tangles, not oscillating superstrings. In fact, the definitions of particles, vacuum and horizons differ completely in the two approaches. In the strand model, in contrast to 'open superstrings', no configuration has ends, and in contrast to open or closed superstrings, strands move in three spatial dimensions, not in nine or ten, and are not related to membranes or supermembranes. In the strand model, no strand is 'bosonic' or heterotic', there is no E(8) or SO(32) gauge group, there are no 'pants diagrams', and the anomaly issue is resolved without higher dimensions. In fact, not a single statement about superstrings is applicable to strands. Ref. 149

* *

Ref. 227 An old trick, known already in France in the nineteenth century, can help preparing for the idea of particle motion in space. Figure 94 shows a special chain that is most easily made with a few dozen key rings. If the ring B is grapped, and the ring A is released, it seems to fall down along the whole chain. If you have never seen it, the result is quite astonishing. It triggers a simple question: Is this old trick a good analogy for the motion of particles in the vacuum? Challenge 153 ny

MOTION THROUGH THE VACUUM – AND THE SPEED OF LIGHT

Up to now, one problem was left open: How can a particle, being a tangle of infinite extension, move through the web of strands that makes up the vacuum? The clarification of particle structures yields the answer:

▷ ***Translational particle motion*** is is due to strand substitution, or 'strand hopping'.

An illustration of translational motion is given in Figure 95. It is easy to picture a photon hopping from strand to strand. For massive particles, the challenge is tougher.

For massive particles, motion is a more complicated process. If a particle were a specific open knot or a specific primary tangle *all the time*, motion would *not* be possible. This impossibility is the main reason that extended models for particles have not be considered before. But in the strand model, mass is due to strand processes occurring at the border of space. In the strand model, the tangles of massive particles change over time: each massive particle is described by an infinite family of tangles. What we call an electron, for example, is the *time-average* of its tangle family.

For the motion of a massive fermion, such as an electron, Figure 95 shows that unwinding processes at the border of space are necessary; these processes have a low probability. In the strand model, this is the reason that massive particles move more slowly than light. For the motion of a W boson, the hopping also requires processes at the border of space.

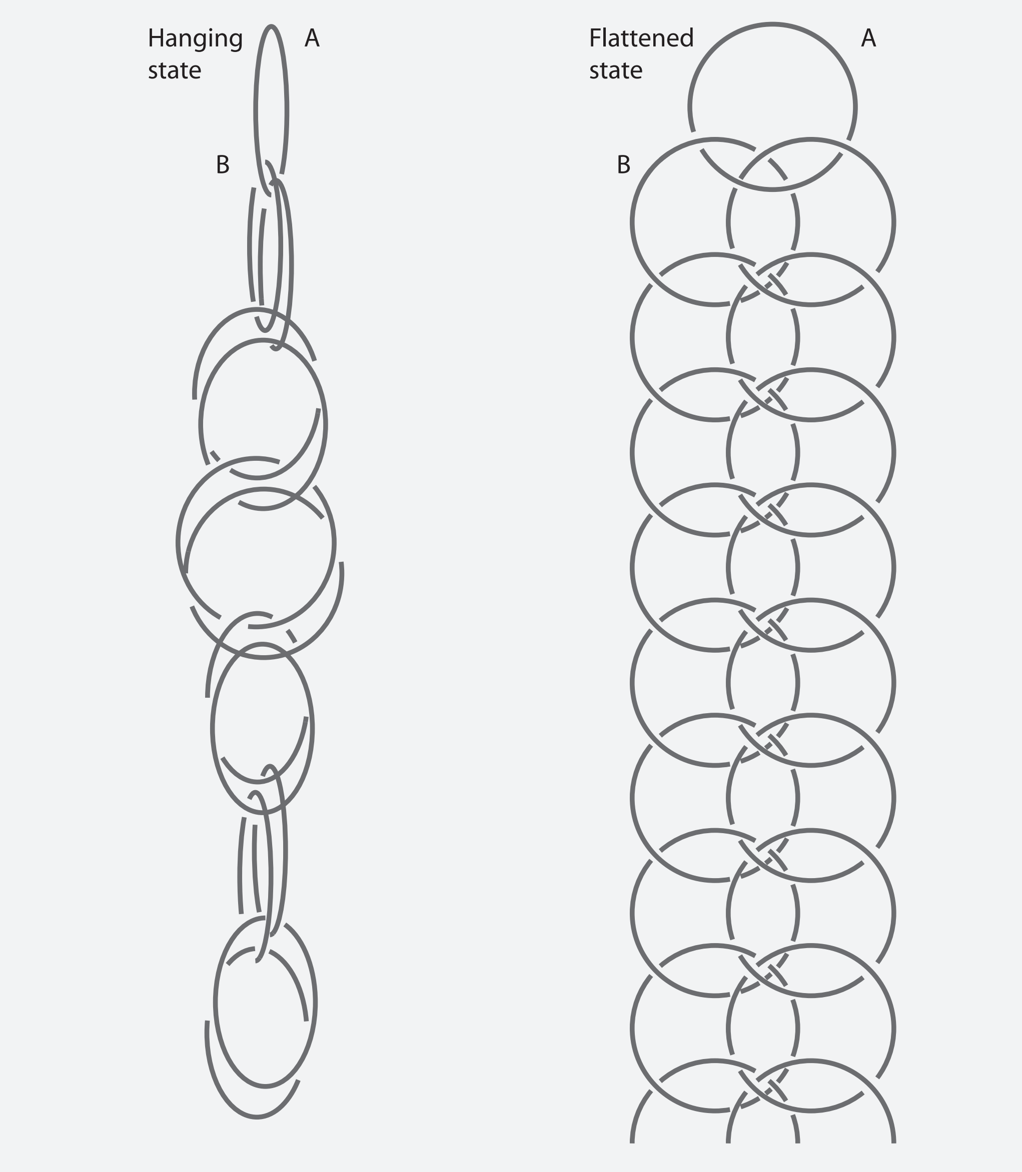


FIGURE 94 A ring chain gives an impression of motion along the chain, when holding ring B while dropping ring A.

In other words, tangles of massive particles *can* move through the vacuum, and this automatically happens more slowly than the motion of photons. The speed of photons is a limit speed for massive particles; special relativity is thus recovered completely.

Summary on millennium issues and predictions about particles

We found that the strand model makes a strong statement: elementary particles can only be made of one, two or three strands. For *one-stranded* particles, the strand model shows

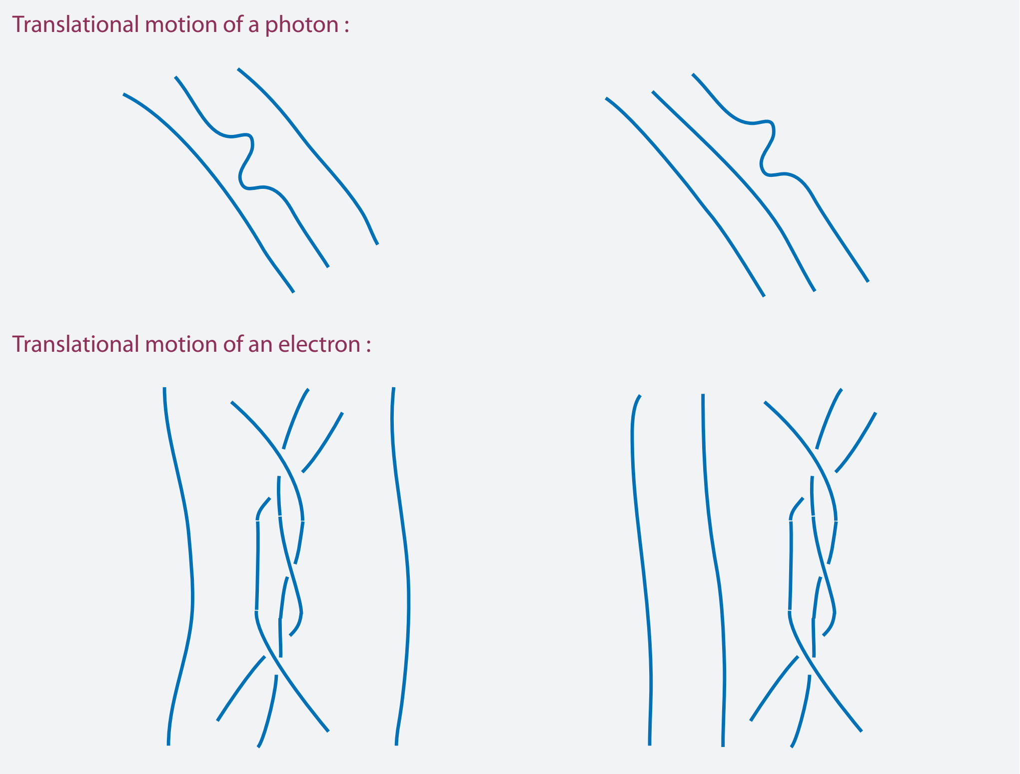


FIGURE 95 Motion of photons and electrons.

that the photon, the W, the Z and the gluons form the full list of spin-1 bosons. For *two-stranded* particles, the strand model shows that there are precisely three generations of quarks. For *three-stranded* elementary particles, the strand model shows that there are three generations of leptons. Neutrinos and antineutrinos differ and are Dirac particles. The strand model thus predicts that the neutrino-less double-beta decay will *not* be observed. Glueballs probably do not to exist.

Each particle is represented by a *family* of tangles. The family members are related through various degrees of knotting and through the leather trick.

The strand model explains the origin of all quantum numbers of the observed elementary particles. Also all predicted quantum numbers for composed particles agree with observations. Therefore, we have also completed the proof that all observables in nature are due to crossing switches. The strand model reproduces the quark model, including all the allowed and all the forbidden hadron states. For mesons and baryons, the strand model predicts the correct mass sequences and quantum numbers. Tetraquarks are predicted to exist. A way to calculate hadron form factors is proposed. Page 152

In the strand model, all tangles are mapped to known particles. The strand model predicts that *no* elementary particles outside the standard model exist, because no tangles are left over. For example, there are no axions, no leptoquarks, and no supersymmetric particles in nature. The strand model predicts the lack of other gauge bosons and other interactions. In particular, the strand model predicts the non-existence of the Higgs boson, for two reasons: spin-0 elementary particles are impossible in the strand model, and so are other elementary particles. In fact, any new elementary particle found in the future would contradict and invalidate the strand model.

The strand model thus shows that the number 3 that appears so regularly in the standard model of particle physics – 3 generations, 3 forces, 3 colours and SU(3) – is, in each case, a consequence of the three-dimensionality of space. In fact, the strand model adds a further, but related number 3 to this list, namely the maximum number of strands in elementary particles.

In simple words, the strand model explains why the known elementary particles exist and why others do not. We have thus settled two further items from the millennium list of open issues. Page 158

Predictions about dark matter

Astrophysical observations show that galaxies and galaxy clusters are surrounded by large amounts of matter that does not radiate. This unknown type of matter is called *dark matter*.

In the strand model, the known particles are the only possible ones. Therefore, the galactic clouds made of dark matter must consist of those particles mentioned up to now. The strand model thus predicts that dark matter is a mixture of the particles of the standard model and black holes. This statement settles a further item from the millennium list of open issues. Page 158

The prediction of a lack of new elementary particles in dark matter is at odds with the most favoured present measurement interpretations, but cannot yet be ruled out. In fact, the prediction provides another hard test of the model: if dark matter is found to be made of non-standard particles, the strand model is wrong.

No discoveries at the LHC – and no science fiction

We can summarize all the results found so far in the following way: general relativity and the standard model of particle physics are both correct descriptions of nature. The strand model provides the basis for both of them. In other words, starting from the relation between Planck units and the crossing switch, we found that *there is nothing to be discovered about nature outside general relativity and the standard model of particle physics*. There is no hidden aspect of nature. The strand model predicts a so-called high-energy *desert*: it predicts the lack of the Higgs boson and of any other new elementary particle. In short, there is no room for discoveries at the Large Hadron Collider in Geneva, nor at the various dark matter searches. In other words, the strand model predicts a lack of any kind of science fiction in this domain.

We have thus shown that the Planck units, via strands, explain almost everything known about motion: in particular, they explain *what* moves and *how* it moves. Only three open issues remain: the masses, mixings and couplings of the elementary particles.

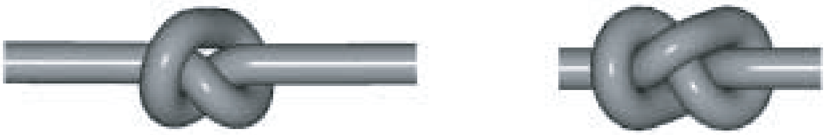


FIGURE 96 A tight open overhand knot and a tight open figure-eight knot (© Piotr Pieranski, from Ref. 228)

THE MASSES OF THE ELEMENTARY PARTICLES

The *mass* describes the inertial and gravitational effects of a body. The strand model must reproduce all mass values observed in nature; if it doesn't, it is wrong. To reproduce the masses of *all* bodies, it is sufficient that the strand model reproduces the measured masses, mixing angles and the couplings of the *elementary* particles. We start with the masses.

In nature, *gravitational mass* is the space curvature induced by a particle. In the strand model, this curvature is due to the fluctuations that originate at a tangle core, especially the tail fluctuations. The value of the mass is strongly influenced by the topology changes that occur in the strand model.

In contrast, *inertial mass* appears in the Dirac equation. In the strand model, inertial mass is is determined by the frequency and wavelength of the rotating phase vector. These quantities are influenced by the type of tangle, by the fluctuations induced by the particle charges, by the topology changes induced by the weak interaction, and in the case of fermions by the average frequency and size of the belt trick. Also these processes are all due to strand fluctuations. For example, tangles with simple structure are expected to have lower mass than tangles with intricate structure. As another example, chiral tangle cores, being sensitive to twisted loops, rotate with more ease than achiral ones (all other properties being comparable), meaning that chiral tangles, i.e., charged particles, have comparatively higher mass; they also fluctuate less easily than achiral tangles, again pointing to higher mass.

In short, both gravitational and inertial particle mass are due to strand fluctuations, more precisely, to tail fluctuations. The strand model thus suggests that gravitational and inertial mass are equal automatically. In particular, the strand model suggests that every mass is surrounded by fluctuating crossing switches whose density decreases with distance and whose number is proportional to the mass itself. This idea leads to universal gravity, as discussed above. Finally, the strand model also suggests that the more complex a tangle is, the higher its mass is. We now look for ways to determine the mass values from the tangle structure. We discuss each particle class separately; we first look at mass ratios, and then at absolute mass values.

Page 249

Boson mass ratios and the weak mixing angle

Mass calculations are especially simple for the W and Z bosons, because in the strand model, they are *clean* systems: each boson is described by a relatively simple tangle family, and W and Z bosons do not need the belt trick to rotate.

We expect that the induced curvature, and thus the gravitational mass, of an elementary boson should be a function of the ropelength of the corresponding tight tangle. *Tight* or *ideal* knots and tangles are those knots or tangles that appear if we imagine strands as being made of a rope of *constant* diameter that is *infinitely flexible* and *infinitely slippery*. Two examples of tight knots are shown in Figure 96. Tight knots and tangles do not exist in everyday life; they are mathematical idealizations. Tight tangles are of interest because they realize the Planck limit of the strand model, if we imagine that each strand has a diameter of one Planck length.

The *ropelength* of a tight *closed* knot is the length of a perfectly felxible and slippery rope of constant diameter required to tie the tight knot. The *ropelength* of an tight *open* knot is the length by which a rope tied into a tight knot is shortened. With some care, ropelength can be also be defined for tangles. In the following, the ropelength is assumed to be measured in units of the rope *diameter*. The ropelength is an obvious choice for measuring the amount by which a tight knot or tangle disturbs the vacuum around it.

It is known from quantum field theory that the masses of W and Z bosons do not change much between Planck energy and low energy, whatever renormalization scheme is used. Ref. 229

In the strand model, the gravitational mass of a spin 1 boson is given by the radius of the disturbance that it induces in the vacuum. For a boson, this radius, and thus the mass, scales as the third root of the ropelength of the corresponding tight knot. The simplest tangle of the W boson is an open overhand knot, and that of a Z boson is an open figure-eight knot, as shown in Figure 96. The corresponding ropelength values for tight tangles are 10.1 and 13.7 rope diameters. The strand model thus predicts a W/Z mass ratio given by the cube root of the ropelength ratio: Ref. 228

$$\frac{m_W}{m_Z} = \left(\frac{L_W}{L_Z}\right)^{1/3} = 0.90 \ . \tag{203}$$

Ref. 218 This value has to be compared with the experimental ratio of 80.4 GeV/91.2 GeV=0.88. The agreement is satisfactory.* In particular, the higher value of the neutral Z boson's mass is reproduced. We will explain below why this approximation works so well. We note that the simple open knots represent W and Z bosons only to a certain approximation, as mentioned above. The strand model predicts that the match between the calculated and the measured ratio m_W/m_Z should improve when higher-order Feynman diagrams, and thus more complicated knot topologies, are taken into account. This is still a subject of research. Page 311 Page 229

The W/Z mass ratio also determines the weak mixing angle θ_w of the weak interaction Lagrangian, through the relation $\cos\theta_w = m_W/m_Z$. The strand model thus predicts the value of the weak mixing angle to the same precision as it predicts the W/Z mass ratio.

---

* If the arguments against the existence of the Higgs should turn out to be wrong, then the strand model might be saved by arguing that the tangle of Figure 83 might effectively have spin 0. This 'tweaked' strand model then would predict *exactly one* Higgs boson. In this case, the ropelength of the Borromean rings, 29.03, would lead to a Higgs mass prediction of 114 GeV. However, this is not a prediction to be taken seriously.

This argument leads to a puzzle: Can you deduce from the strand model how the W/Z mass ratio changes with energy? Challenge 154 ny

The inertial mass of the W and Z bosons can also be explored. In quantum theory, the inertial mass relates the wavelength and the frequency of the wave function. In the strand model, a quantum particle that moves through vacuum is a tangle that rotates while advancing. Page 174 The frequency and the wavelength of the helix thus generated determine the inertial mass. The process is analogous to the motion of a body moving at constant speed in a viscous fluid at small Reynolds numbers. Despite the appearance of friction, the analogy is possible. If a small body of general shape is pulled through a viscous fluid by a constant force, such as gravity, it follows a *helical* path. Ref. 230 This analogy implies that, for spin 1 particles, the frequency is mainly determined by the radius of the small body. The strand model thus predicts that the inertial mass of the W and Z bosons, like the gravitational mass, is proportional to their radius, as required by consistency.

In summary, the strand model predicts a W/Z mass ratio and thus a weak mixing angle close to the observed value. We also recall that, in the strand model, particle and antiparticle masses are always equal.

Quark mass ratios

The strand model makes several predictions about quark masses. First of all, the simplest quark tangles are clearly less tangled than the tangles of the W and Z boson. The model thus predicts that light quarks are less massive than the W and Z bosons, as is observed. The quark masses are also predicted to be the same for each possible colour charge, as is observed.

Furthermore, the progression in ropelength of the tight basic tangles for the six quarks suggests a progression in their masses. This is observed, though with the exception of the up quark mass. For this exceptional case, mixing effects (due to the leather trick) are expected to play a role, as argued below.

For each quark number q, the quark mass will be an average over braids with q, $q+6$, $q+12$... crossings, where the period 6 is due to the leather trick. Each tight braid has a certain ropelength L. Its mass will be given by the frequency of the belt trick. The belt trick frequency will be an exponentially small function of the ropelength; we thus expect a general mass relation of the type

$$m \sim \mathrm{e}^{aL} \qquad (204)$$

where a is a number of order 1. We note directly that such a relation promises agreement with the observed ratios among quark masses. Ref. 231 More detailed ropelength calculations show that the ropelength increases approximately linearly with q, as expected from general knot theoretic arguments. The calculated ropelength values suggest that a has an effective value in the range between 0.4 and 0.9.

The strand model predicts that quark masses result from a combination of the effects of ropelength and of the leather trick. The strand model thus predicts

$$m_q \sim p_q \mathrm{e}^{aL_q} + p_{q+6} \mathrm{e}^{aL_{q+6}} + \ldots \qquad (205)$$

where p_q is the probability of the tangle with q crossings. So far however, the values of these probabilities are unknown.

In the approximation that the first probability is 1 and all the others 0, we get only Ref. 231 a very poor match between observed and calculated masses. The only encouraging aspect is that this approximation could provide the underlying reason for previous speculations on approximately *fixed ratios* between up-type quark masses and *fixed mass ratios* Ref. 232 between down-type quark masses.

The probabilities due to the leather trick play an important role for mass values. We note that the strand model predicts a very small bare mass, i.e., Planck energy mass, for the down quark. In fact, the down quark mass at Planck energy would vanish if there were no leather trick. However, in nature, the down mass is observed to be larger than the up mass. The leather trick has the potential to explain the exceptionally large mass of the Ref. 231 down quark; the added ropelength of six additional crossings is about 21; the resulting factor e^{a21} is in the range between a million and a few thousand millions. This large factor could compensate a small probability for the tangle with 6 additional crossings, thus leading to a down quark mass that is higher than suggested by the simplicity of its most basic tangle.

The probabilities in the quark mass formula (205) are also expected to explain the deviations from a simple exponential increase of quark mass with quark generation. In fact, the precise determination of the average *shape* of quark tangles promises to yield *different* mass ratios among up-type and among down-type quarks. All these issues are still subject of research.

LEPTON MASS RATIOS

Mass calculations for leptons are involved. Each lepton has a large family of associated tangles: there is a simplest tangle, there are the tangles that appear through repeated application of the leather trick, and there are the tangles that appear through higher-order propagators due to the weak interaction. Despite this large families, some results can be deduced from the simplest lepton tangles alone, disregarding the higher-order complications.

First of all, the simplest lepton tangles are much less knotted than the simplest 'broken' W and Z boson tangles. The strand model thus suggest that the masses of the leptons are much smaller than those of the W and Z boson, as is observed.

But we can say more. For the electron neutrino, the fundamental tangle is almost unknotted. The strand model thus suggest an almost vanishing mass for the electron neutrino, were it not for the leather trick and the weak interaction effects. On the other hand, the electron is predicted to be more massive than its neutrino due to its electric charge.

The progression in ropelength of the tight versions of the simplest tangles for the six leptons suggests a progression in their masses with generation, if equal charges are compared.

For each lepton generation, the lepton mass will be an average over braids with crossing numbers l, $l+6$, $l+12$, etc. For each lepton braid with l crossings, knot theory predicts a ropelength L that increases roughly proportionally to l. Its mass will again be given by the frequency of the belt trick and the probability of the leather trick. We thus expect a general relation of the type

$$m_l \sim p_l \mathrm{e}^{bL_l} + p_{l+6} \mathrm{e}^{bL_{l+6}} + \ldots \tag{206}$$

where b is a number of order 1. Such a relation is in general agreement with the observed ratios between lepton masses.

More precise predictions of lepton mass ratios will become possible with the help of numerical simulations. We discuss this possibility below. We note again that the mass generation mechanism of the strand model contradicts several other proposals in the research literature. In particular, it contradicts the Higgs mechanism and, for neutrinos, the see-saw mechanism. The strand model is more on the line of the idea of conformal symmetry breaking. Ref. 233

THE MASS HIERARCHY: MASS RATIOS ACROSS PARTICLE FAMILIES

Boson tangles are made of one strand, quarks tangles of two, and leptons tangles of three strands. Each strand reduces the probability of the belt trick, thus effectively reducing the mass. As a result, the strand model predicts that the mass ratios for the *least massive* members of the three particle families, namely massive bosons, quarks and leptons, – if the leather trick is neglected – follow

$$\frac{m_b}{m_q} \approx \frac{m_q}{m_l} \approx F \,, \tag{207}$$

where F is a constant. The observed ratios are $1.6 \cdot 10^5$ and around 10^6; the second value is not precisely known, because the masses of the neutrinos are not precisely known yet. The agreement with the prediction is satisfactory, especially since this relation is only expected to be roughly correct, since it is modified by the leather trick. In short, the strand model suggests the existence of a mass hierarchy.

In summary, the strand model predicts that fermion mass values, before the leather trick is taken into account, behave as

$$m_f \approx B\, F^{-n}\, \mathrm{e}^{c(n)L_f} \tag{208}$$

where n is the strand number, L the ropelength of the tight tangle, $c(n)$ is a number of order one that was called a or b above, F a number of the order of 10^5, and B a value of the order of 10^{-17}. The observed fermion masses roughly obey this relation. The strand model also predicts that the effects of the leather trick will produce deviations from this simple relation. The factor F is expected to follow from calculations on the probability of the belt trick. This calculation is still subject of research.

One issue remains open: What is the origin of the proportionality factor B that yields the absolute values of particle masses?

PREDICTIONS ABOUT ABSOLUTE MASS VALUES

First of all, the strand model predicts mass sequences and mass ratios that corroborate or at least do not contradict observations. The strand model also predicts that particle masses do not depend on the age of the universe, in agreement with all observations so far. The strand model also predicts that particles and antiparticles have the same mass, again in agreement with all measurements so far.

To determine *absolute* mass values, we need, for all particles, the probability for strand overcrossing at the border of space. It is clear that the probability is low, since, sloppily speaking, the border of space is far away. This implies that all particle masses are predicted to be much smaller than the Planck mass, as is observed.

The strand model thus reduces the calculation of absolute particle masses to the calculation of a single process: the overcrossing of strands at the border of space. We note that in the past, various researchers have reached the conclusion that all elementary particle masses should be due to a single process or energy scale. For example, the breaking of conformal symmetry has always been a candidate for the associated process. Also the Higgs mechanism is a proposed common origin for all particle masses. Experiments in 2010 at the Large Hadron Collider in Geneva should decide which approach is the correct description of nature. Ref. 233, Ref. 234

In other terms, in the strand model, *absolute* mass values seem *not* to be purely geometrical quantities that can be deduced from tangle shapes. Particle masses are to a large extent dynamical, and given by the probability of strand overcrossing at the border of space. We thus find again that absolute mass values are due to strand fluctuations, and that overcrossing is the most important among the fluctuations. If we prefer, masses are due to the relative probability of the various tangles in the family of tangles that represent a particle.

In particular, we expect the strand fluctuations to be similar for particles with the *same* number of strands. This explains why determining the Z/W boson mass *ratio* or the fermion mass *ratios* is possible with some precision: for these mass ratios, the overcrossing processes cancel out.

To determine the absolute particle mass, we need to determine the ratio between the particle mass and the Planck mass. In other words, we have to determine the ratio between the overcrossing probability for the particle and the overcrossing probability for a Planck mass. In the strand model, a Planck mass seems to be an unknot. However, its overcrossing probability is unclear. In short, finding an analytical approximation for absolute particle masses is still an open issue.

OPEN ISSUE: CALCULATE MASSES AB INITIO

Calculating absolute particle masses from tangle fluctuations, either numerically or with an analytical approximation, will allow the final check of the statements in this section. The strand model predicts that the result, which will determine the factors p, F and B just mentioned, will match experiments. Challenge 155 ny

* *

Ref. 235 Challenge 156 ny Will the recently discovered, almost linear relation between ropelength and crossing number help in these calculations?

Summary on particle masses and millennium issues

The strand model provides estimates for particle mass ratios and for the weak mixing angle. Mathematically, masses are found to be geometric properties of specific tangle families. Masses are thus predicted to be constant in time, as is observed. Mass sequences and first rough estimates of mass ratios agree with the experimental data. The strand model also promises to calculate absolute mass values. Future numerical simulations will allow either to improve the match with observations or to refute the strand model.

Page 158

In the millennium list of open issues, we have thus shown how to settle several further items – though we have not settled them completely yet. We now continue with the investigation of particle mixing.

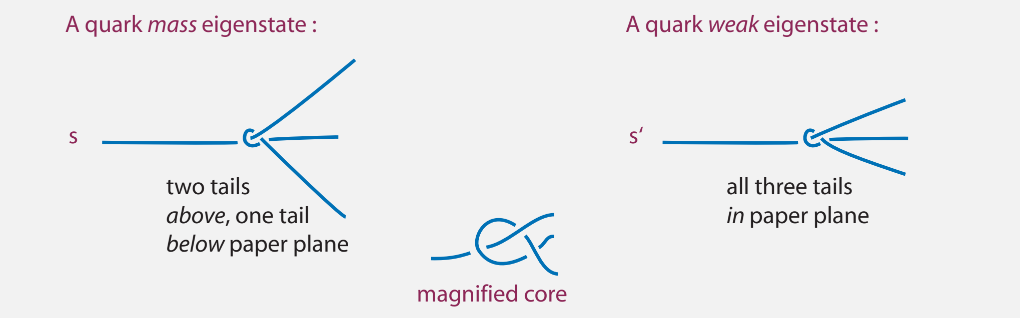


FIGURE 97 Overcrossing leads to quark mixing: mass eigenstates and weak eigenstates differ.

MIXING ANGLES

In nature the *mass* eigenstates for fermions differ from their *weak* eigenstates: quarks mix; so do neutrinos. The mixing is described by the so-called *mixing matrices*. If the strand model does not reproduce this observation and the measured values, it is wrong.

QUARK MIXING

In nature, the quark mass eigenstates and their weak eigenstates differ. This effect, discovered in 1963 by Nicola Cabibbo, is called *quark mixing*. The values of the elements of Ref. 218 the quark mixing matrix have been measured in detail, and experiments to increase the precision are still under way.

Vol. V, page 181 The *mixing matrix* is defined by

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = (V_{ij}) \begin{pmatrix} d \\ s \\ b \end{pmatrix} . \qquad (209)$$

where, by convention, the states of the +2/3 quarks u, c and t are unmixed. Unprimed quarks names represent strong eigenstates, primed quark names represent weak eigen-Ref. 218 states. In its standard parametrization, the mixing matrix reads

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}\mathrm{e}^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}\mathrm{e}^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}\mathrm{e}^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}\mathrm{e}^{i\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}\mathrm{e}^{i\delta_{13}} & c_{23}c_{13} \end{pmatrix} \qquad (210)$$

where $c_{ij} = \cos\theta_{ij}$, $s_{ij} = \sin\theta_{ij}$ and i and j label the generation ($1 \leqslant i, j \leqslant 3$). The mixing matrix thus contains three mixing angles, θ_{12}, θ_{23} and θ_{13}, and one phase, δ_{13}. In the limit $\theta_{23} = \theta_{13} = 0$, i.e., when only two generations mix, the only remaining parameter is the angle θ_{12}, called the *Cabibbo angle*. The phase δ_{13}, lying between 0 and 2π, is different from zero in nature, and expresses the fact that CP invariance is violated in the case of the weak interactions. It appears in the third column and shows that CP violation is related to the existence of (at least) three generations.

The present 90 % confidence values for the measured *magnitude* of the complex quark mixing matrix elements are Ref. 218

$$|V| = \begin{pmatrix} 0.97419(22) & 0.2257(10) & 0.00359(16) \\ 0.2256(10) & 0.97334(23) & 0.0415(11) \\ 0.00874(37) & 0.0407(10) & 0.999133(44) \end{pmatrix} . \qquad (211)$$

Within experimental errors, the matrix V is unitary. A huge amount of experimental work lies behind this short summary. The data has been collected over many years, in numerous experiments, by thousands of people. Nevertheless, this short summary represents all the data that any unified description has to reproduce.

In the standard model of particle physics, the quark mixing matrix is usually seen as due to the coupling between the vacuum expectation value of the Higgs field and the left-handed quark doublets or the right handed quark singlets. In contrast, in the strand model, the Higgs field does not exist, and its role as mass generator and unitarity maintainer is replaced by the process of strand overcrossings at the border of space. In the strand model, overcrossing is related to the weak interaction. Because the various quarks are differently braided rational tangles, overcrossing can reduce or increase the crossings in a quark tangle, and thus change quark flavours. We thus deduce that quark mixing is an automatic result of the strand model and related to the weak interaction. We also deduce that quark mixing is due to the *same* process that generates quark masses, as expected. But we can say more. Page 229

In the strand model, the mass (and colour) eigenstate is the shape in which colour symmetry is manifest and in which particle position is defined. The *mass eigenstates* of quarks correspond to tangles whose three colour-tails point in three directions that are equally distributed in space. The shape in which the tails point in three, equally spaced directions is the shape that makes the SU(3) representation under core slides manifest.

In contrast, the *weak eigenstates* are those shapes that makes the SU(2) behaviour of core pokes manifest. For a quark weak eigenstate, the shape seems to be that of a tangles whose three tails lie in a plane. The two types of eigenstates are illustrated in Figure 97.

We call transformation from a mass eigenstate to a weak eigenstate or back *tail shifting*. Tail shifting is a deformation: the tails are rotated as a whole and shifted. On the other hand, tail shifting can also lead to untangling of the braid; in other words, tail shifting can lead to strand overcrossing and thus can transform quark flavours. Tail shifting can be seen as a *partial* overcrossing; as such, it is due to the weak interaction.

Tail shifting, both with or without strand overcrossing at the border of space, is a generalized deformation. As such, it is described by a unitary operator. The first result from the strand model is thus that the quark mixing matrix is unitary. This is indeed observed within experimental errors. Page 193 Ref. 218

For quarks in mass eigenstates, and thus in particular for quarks bound in hadrons, overcrossing is a process with small probability. As a consequence, the quark mixing matrix will have its highest elements on the diagonal. This is indeed observed.

The strand model predicts that quark mixing will be higher between neighbouring generations, such as 1 and 2, than between distant generations, such as 1 and 3. This is also observed.

The connection between mixing and quark mass also implies that the 1–2 mixing is larger than the 1–3 mixing, as is observed.

Finally, the strand model predicts that the numerical values in the quark mixing matrix can be deduced from the difference between the geometries of the two kinds of tangles shown in Figure 97. Performing a precise calculation of mixing angles is still a subject of research.

A CHALLENGE

Ref. 236 Can you deduce the approximate expression

$$\tan\theta_{\mathrm{u\,mix}} = \sqrt{\frac{m_{\mathrm{u}}}{m_{\mathrm{c}}}} \qquad (212)$$

Challenge 157 ny for the mixing of the up quark from the strand model?

CP-VIOLATING PHASE FOR QUARKS

The CP violating phase δ_{13} for quarks is usually expressed with the *Jarlskog invariant*, defined as $J = \sin\theta_{12}\sin\theta_{13}\sin\theta_{23}^2\cos\theta_{12}\cos\theta_{13}\cos\theta_{23}\sin\delta_{13}$. This involved expression is independent of the definition of the phase angles and was discovered by Cecilia Jarlskog, an important Swedish particle physicist. Its measured value is $J = 3.05(20)\cdot 10^{-5}$. Ref. 218

Since the strand model predicts exactly three quark generations, the quark model predicts the possibility of CP violation. We showed already in the section on mesons that this possibility is indeed realized. Figure 87 showed that with a combination of overcrossings, K^0 and $\overline{K}^0$ mesons will mix, and that the same happens with certain other neutral mesons. The strand model also predicts that the effect is small, but non-negligible, as is observed. Page 293 Page 291

The strand model thus predicts that the quark mixing matrix has a non-vanishing CP-violating phase. The value of this phase is predicted to follow from the geometry of the quark tangles, if their fluctuations are properly accounted for. This topic is still a subject of research.

NEUTRINO MIXING

The observation, in 1998, of neutrino mixing is comparably recent in the history of particle physics, even though the Italian-Soviet physicist Bruno Pontecorvo predicted the effect already in 1957. Again, the observation of neutrino mixing implies that the mass eigenstates and the weak eigenstates differ. The values of the mixing matrix elements are only known with limited accuracy so far; it is definitely known, however, that the mixing among the three neutrino states is strong, in contrast to the situation for quarks. Neutrino masses are known to be positive. So far, experiments only yield values of the order of 1 ± 1 eV. Ref. 237 Ref. 218

In the strand model, the mass eigenstates correspond to tangles whose tails point along the three coordinate axes. In contrast, the weak eigenstates again correspond to tangles whose tails lie in a plane. The two kinds of eigenstates are illustrated in Figure 98.

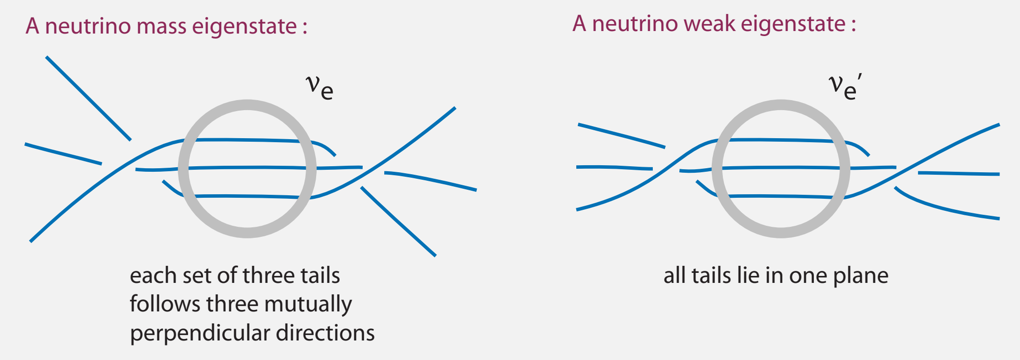


FIGURE 98 Overcrossing leads to neutrino mixing: mass eigenstates and weak eigenstates differ.

Again, the transition between the two eigenstates is due to tail shifting, a special kind of deformation.

We thus deduce that neutrino mixing, like quark mixing, is an automatic result of the strand model and is related to the weak interaction. Given that the neutrino masses are low and similar, and that neutrinos do not form composites, the strand model predicts that the mixing values are large. This is a direct consequence of the leather trick, which in the case of similar masses, mixes neutrino tangles with 0, 2, 4, 6, 8, 10 etc. crossings and thus leads to large mixings between *all* generations, and not only between neighbouring generations. In the strand model, the large degree of neutrino mixing is thus seen as a consequence of their low and similar masses, and of their existence as free particles.

The strand model predicts a *unitary* mixing matrix for neutrinos. The strand model also predicts that the geometry of the neutrino tangles and their fluctuations will allow to calculate the mixing angles. More precise predictions are still subject of research.

CP-violation in neutrinos

The strand model predicts that the three neutrinos are massive Dirac particles. This has not yet been verified by experiment. The strand model thus predicts that the neutrino mixing matrix has *only one* CP-violating phase. (It would have three such phases if neutrinos were Majorana particles.) The value of this phase is predicted to follow from the neutrino tangles and a proper accounting of their fluctuations. Also this topic is still a subject of research. On the other hand, it is unclear when the value of the CP-violating phase will be measured. This seems the hardest open challenge of experimental particle physics – provided that the strand model is correct.

The mechanism of CP violation has important consequences in cosmology, in particular for the matter–antimatter asymmetry. Since the strand model predicts the absence of the see-saw mechanism, the strand model rules out leptogenesis, an idea invented to explain the lack of antimatter in the universe. The strand model is more on the line with electroweak baryogenesis.

Ref. 238

Ref. 239

OPEN CHALLENGE: CALCULATE MIXING ANGLES AND PHASES AB INITIO

Calculating the mixing angles and phases ab initio, using the statistical distribution of strand fluctuations, is possible in various ways. In particular, it is interesting to find the relation between the probability for a tail shift and for an overcrossing. This will allow to check the statements of this section. Challenge 158 ny

SUMMARY ON MIXING ANGLES AND THE MILLENNIUM LIST

We have shown that tangles of strands predict non-zero mixing angles for quarks and neutrinos, as well as CP-violation in both cases. The strand model also predicts that the mixing angles of quarks and neutrinos can be calculated from strand fluctuations. The general properties of the mixing matrices agree with the experimental data as known today. The strand model also predicts that all angles and phases are constant in time. Finally, the strand model rules out leptogenesis.

We have thus partly settled four further items from the millennium list of open issues. Page 158 Future numerical calculations will allow either to improve the checks or to refute the strand model.

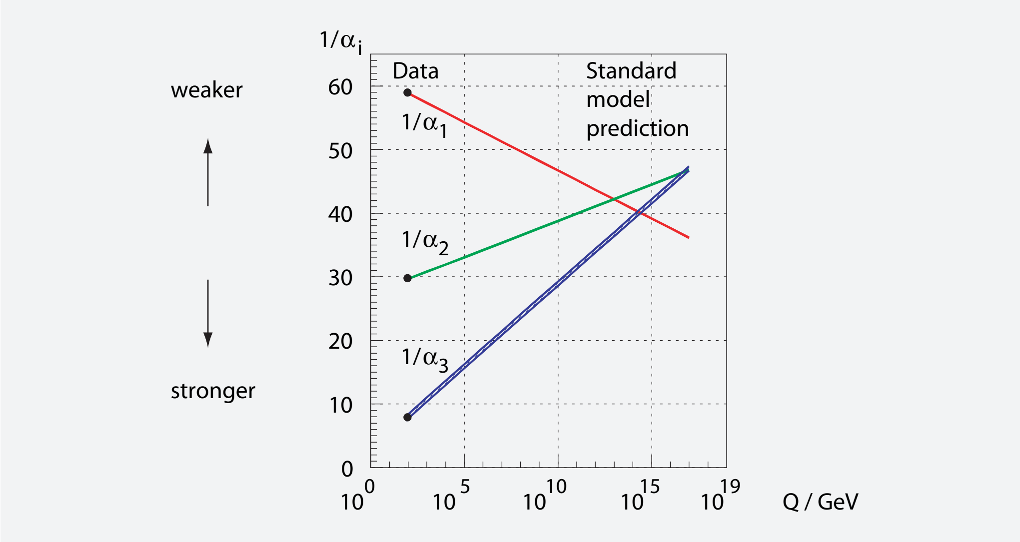


FIGURE 99 The behaviour of the three coupling constants squared with energy, as predicted by the standard model of particle physics; the graph shows the constants $\alpha_1 = \frac{5}{3}\alpha/\cos^2\theta_W$ for the electromagnetic interaction (the factor 5/3 is important in grand unification), $\alpha_2 = \alpha/\sin^2\theta_W$ for the weak interaction, and $\alpha_3 = \alpha_s$ for the strong coupling constant. The three points are the data points for the highest energies measured so far; at lower energies, data and calculation matchwithin experimental errors (courtesy Wim de Boer).

COUPLING CONSTANTS AND UNIFICATION

In nature, the strength of a gauge interaction is described by its coupling constant. The coupling constant gives the probability with which a charge emits a virtual gauge boson. There are three coupling constants, one for the electromagnetic, one for the weak and one for the strong interaction. The three coupling constants depend on energy. The known data and the prediction of the change with energy predicted by the standard model of particle physics are shown in Figure 99. At the lowest possible energy (0.511 MeV), the fine structure constant, i.e., the electromagnetic coupling constant, has the well-known value 1/137.035 999 1(1). If the strand model cannot reproduce the data, it is wrong.

In the strand model, interactions are due to shape changes of tangle cores. Given a tangle core, the following strand fluctuations can occur:

— *Small* changes of core shape do not produce any crossing switch. Small shape fluctuations thus have no physical significance: for an observer, they leave all observables unchanged.

— *Twist* shape changes of a strand segment in the core produce an electric field, if the particle is charged. More precisely, the electric field is the difference between the average number p_{tr} of right twists and the average number p_{tl} of inverse, left twists that a particle tangle produces per unit time.

— *Poke* shape changes of a strand segment in the core produce a weak interaction field. More precisely, the weak field is the asymmetry among the probabilities $p_{\mathrm{p}x}$, $p_{\mathrm{p}y}$ and $p_{\mathrm{p}z}$ of the three fundamental poke types and their inverses.

— *Slide* shape changes of a strand segment in the core produce a colour field, if the particle has colour. More precisely, the colour field is the asymmetry among the probabilities p_{s1} to p_{s8} of the eight fundamental slide types and their inverses.

In the strand model, the fluctuation probabilities for each Reidemeister move – twist, poke or slide – determine the coupling constants. We thus need to determine these probability values. Fortunately, a few conclusions can be deduced directly, without any detailed calculation.

First of all, by relating coupling constants to shape fluctuation probabilities, the strand model predicts coupling constants to be *positive* numbers and *smaller than 1* for all energies. This is indeed observed. A still stricter bound for coupling constants can also be deduced. The sum of all possible fluctuations for a particular tangle has unit probability. We thus have

$$1 = p_{\text{small}} + p_{\text{tr}} + p_{\text{tl}} + p_{\text{p}x} + p_{\text{p}y} + p_{\text{p}z} + p_{\text{p}-x} + p_{\text{p}-y} + p_{\text{p}-z} + \sum_{g=1}^{8} (p_{\text{s}g} + p_{\text{s}-g}) \,. \qquad (213)$$

The strand model thus predicts that the *sum* of the three coupling constants must be strictly smaller than 1 for every energy value. This is easily checked, both with the data and with the prediction of quantum field theory. In quantum field theory, the (modified) *square* of the three coupling constants is given, as a function of energy, in the popular graph shown in Figure 99. (In this popular graph, the electromagnetic coupling is traditionally multiplied by $5/(3\cos^2\theta_W)$, in order to test grand unification.) The graph allows to deduce that the sum of the three unmodified couplings is indeed smaller than 1 for all energy values, as predicted by the strand model.

The strand model also predicts that the three coupling constants are related by small numbers, as the corresponding fluctuations differ only in the number of involved strands. This is also observed, as Figure 99 shows – especially if we remember that the couplings are, apart from the mentioned factors, the square roots of the values shown in the graph.

The strand model further predicts that the coupling constants are independent of time, and that in particular, they do not depend on the age of the universe. This is also observed, despite occasional claims to the contrary. Ref. 240

In summary, the strand model implies, like quantum field theory, that coupling constants are probabilities. The obvious consequences that the coupling constants must be smaller than one and sum to a number smaller than one are valid both in quantum field theory and in the strand model. Despite the agreement with experiment however, we have not deduced anything new so far.

In fact, one new point is made by the strand model. Each gauge interaction is due to a different Reidemeister move; but given a specific core deformation, a different observer will classify it into a different Reidemeister class. Indeed, each Reidemeister move can be realized by the deformation of one single strand. The strand model thus provides *unification* of the interactions. This result is new. At energies measurable in the laboratory, however, the three moves *can* be distinguished, and the three interactions differ. Challenge 159 e

Predictions for calculations of coupling constants

The strand model predicts that the calculation of the three coupling constants is a problem of tangle geometry. This it can be approached either *analytically* or with *computer simulations*, at each energy scale. The calculations need to determine the probabilities of the corresponding Reidemeister fluctuations. If the results do not agree with the experimental values, the strand model is false. We note that there is no freedom to tweak the calculations towards the known experimental results.

The strand model predicts that all tangles with unit electric charge have the *same* coupling, and thus the same probability for preferred fluctuations of twists, i.e., of Reidemeister I moves. For example, the preferred probabilities must be identical for the positron and the proton. *The twist probabilities are thus predicted to be quantized and to classify tangles into equivalence classes defined by their chirality.*

So far, there do not seem to exist any analytical tools that permit the calculation of shape deformation probabilities. At present, computer calculations seem to be the only possible choice. Of all existing software programs, the most adapted to calculating fluctuation probabilities are the programs that simulate the dynamics of knotted polymers; but also the programs that simulate the dynamics of cosmic strings or the dynamics of helium vortices are candidates. The biggest difficulty, apart from a large computer time, is the correct specification of the shape fluctuation distribution at each energy scale. Challenge 160 r

The prediction about electric charge can be extended to the nuclear interactions. *The strand model predicts quantized tangle equivalence classes for the weak charge and for the colour charge.* If any of these predictions are found to be incorrect, the strand model is false. However, there are several hints that these predictions are correct.

Predictions on the quantization of charges and on the fine structure constant

In nature, electric, weak and strong charge are *quantized*. No experiment has ever found even the smallest deviation from charge quantization. *All charges in nature are integer multiples of a smallest charge unit.* Specifically, the electric charge of every particle is found to be an integer multiple of the positron electric charge, divided by three. We call the integer, divided by three, the *electric charge quantum number*. The electromagnetic coupling of the positron is 1/11.706 237 615(4) at low energy, i.e., at 0.51 MeV. This value of the electric charge unit is the square root of the famous *fine structure constant* 1/137.035 999 1(1). Quantum electrodynamics also predicts the precise change with energy of this unit; the experiments performed so far, up to over 100 GeV, agree with this prediction. In particular, quantum electrodynamics predicts a change that, when extrapolated to the Planck energy, would yield a charge unit value of 1/10.2(1). If the strand model does not reproduce these facts, it is wrong.

We thus need to understand, using the strand model, the quantization, i.e., the integer multiples, of the electric charge on one hand, and the mysterious value of the charge unit on the other hand.

In the strand model, electric charge is related to the *chirality* of a tangle. Only chiral tangles are electrically charged. The strand model thus implies that a topological quantity for tangles – defined for each tangle in the tangle family corresponding to a specific elementary particle – must represent electric charge. Which quantity could this be?

The usual topological quantity to determine chirality of knots and tangles is the topological writhe. To determine it, we draw a minimal projection, i.e., a two-dimensional knot or tangle diagram with the smallest number of crossings possible. We then count the right-handed crossings and subtract the number of left-handed crossings. This difference is the topological writhe.

- The topological writhe of the open trefoil is +3 or −3, depending on which mirror image we look at; the topological writhe of the open figure-eight knot vanishes. The topological withe of any unknotted strand also vanishes. In this way, if we define the electric charge quantum number as *one third* of the topological writhe, we recover the correct electric charge quantum number of all gauge bosons.
- Page 280 The tangles of the quarks show that if we define the electric charge quantum number as *one third* of the topological writhe, we recover the correct electric charge quantum number of all quarks.* We note that the leather trick does not change this result.
- Page 285 The tangles of the leptons show that if we define the electric charge quantum number as the topological writhe of the *centre region*, we recover the correct electric charge quantum number of all leptons. Again, the leather trick does not change this result.

In other terms, electric charge quantum number can be reproduced with help of the topological writhe.

Let us sum up. In nature, electric charge is quantized. The strand model describes charged particles with the help of fluctuating alternating tangles, and charge quantization is a topological effect that results because all particles are made of strands. In particular, the *electric charge quantum number* behaves like topological writhe: it is quantized, has two possible signs, vanishes for achiral tangles, and is a topological invariant.

Two unclear points remain: Since every particle is described by a tangle family with an infinite number of members, how is the charge/topological writhe of the other tangle family members accounted for? And why is the charge definition different for leptons? We skip these questions for the time being and continue with the hard problem, namely to understand the value of the charge unit.

Estimating the fine structure constant

In the strand model, the (square root of the) fine structure constant is the probability for the emission of twists by a fluctuating chiral tangle. *The strand model predicts that the fine structure constant can be determined by determining the probability of twists in the random tangle shapes of a given particle.* In other words, the strand model predicts that the probability of the first Reidemeister move in chiral particle tangles is quantized. This probability is predicted to be an integer multiple of a unit that is common to all tangles; and this coupling unit is further predicted to be the fine structure constant. Let us check this prediction.

A check for the existence of a coupling unit requires the calculation of twist emission probabilities for each chiral particle tangle. The strand model is only correct if all particles with the same electric charge yield the same twist emission probability. A simple estimate yields the result that numerical simulations of random shapes require a large

---

* This implies that the writhe of quarks – must be one third or two thirds of the W charge.

amount of computer time, due to the large number of configurations that must be explored. Can the twist emission probability be estimated with a simpler calculation? Can we check whether it is quantized? There are a number of possibilities.

A first, simple exploration of the fine structure constant is based on the *conjecture* that the twist emission probability is proportional, at least approximately, to the 3d writhe of Page 345 a tangle. The *3d writhe* is the three-dimensional average of the topological writhe.

To determine 3d writhe, we imagine to observe a knot or tangle from a random direction and draw the projection, i.e., the two-dimensional tangle diagram. We then count the right-handed crossings and subtract the number of left-handed crossings. If we average this difference over all possible observation directions, we get the 3d writhe of that knot or tangle. In short, the 3d writhe is a three-dimensional measure of chirality. The important point is that its value *depends on the shape* of the tangle. To get a physically useful observable for the strand model, we must then also average over all possible tangle shapes (for a given wave function).

3d writhe is a natural candidate for full electric charge, which is the product of the fine structure constant and the electric charge quantum number. 3d writhe is a generalization of topological writhe, and topological writhe behaves, as we just saw, like the electric charge quantum number. Therefore, the conjecture can also be called an approximation.

Recent research on knot and tangle shape averaging has produced some important results about 3d writhe and other properties. Researchers discovered that several average properties of *random* knots and tangles correlate with properties of *tight* knots and Page 306 Ref. 240 tangles. Two examples of tight tangles are shown in Figure 96.

To clarify the correlation, we take a set of fluctuating or random tangles, all with the same topology; these tangles show fluctuating or random values of the 3d writhe. Nevertheless, the *average* 3d writhe for all tangle shapes is found numerically to be indis- Ref. 241 tinguishable from the 3d writhe of the *tight* tangle. This has been confirmed in various studies to a precision of a few per cent. For example, both the average 3d writhe of any *random* achiral knot and the 3d writhe of the corresponding *tight* achiral knot vanishes. This equality has been checked numerically for knots with minimal crossing numbers below 10.

The picture is completed by a second, interesting result of modern knot research. In the years between 1996 and 1998 it was discovered that the 3d writhe for closed alternat- Ref. 242 ing *tight* knots is *quasi-quantized*. More specifically, many different closed knots share almost the same 3d writhe value; in particular, to within 1%, the 3d writhe of all (small) closed alternating knots is the multiple of a 'writhe quantum', with the value 4/7.*

Let us sum up. If the electric charge unit, i.e., the twist emission probability of simply charged particles, would be given by 3d writhe, then the charge unit could be related to the quasi-quantum of the 3d writhe.** Of course, one would need to check first that all mentioned results also hold for tangles, and not only for knots.

---

* Why is the writhe of alternating tight knots quasi-quantized? No simple argument is known yet.

** We note that the issue of quasi-quantization – instead of an exact quantization – could be circumvented in the following way. In the strand model, at low energy, the strands effectively have negligible thickness. As a result, the average writhe of randomly shaped tangles could deviate somewhat from the writhe of tight tangles. This could imply that the average writhe of loose, randomly shaped tangles might be exactly quantized, whereas the writhe of tight tangles remains only approximately quantized.

Alas, the known value of the writhe quasi-quantum, 4/7, does *not* help us to deduce the value of the fine structure constant, because in this conjecture on the origin of the fine structure constant, the writhe quasi-quantum could be multiplied by any numerical factor to yield the particle coupling. Nevertheless we are left with a fascinating conjecture that suggests a natural way in which electric charge is quantized in nature, and thus explains why the fine structure constant exists.

Towards an estimate of the fine structure constant

We mentioned that each coupling constant gives the average probability of virtual boson emission. In the case of electromagnetism, it might be that the *total torsion* of a tight tangle is a more accurate measure for the probability of twist emission. This option is under exploration. In any case, estimating the twist emission probability without computer simulations is not a problem that is easy to solve without deeper insight into tangle shapes.

A slightly different view of the coupling constants also seems promising. The fine structure constant can also be seen as the average value by which the phase of a wave function changes when a photon is absorbed (or emitted). The strand version of this view offers an interesting approach for calculations.

In this approach, the simplest situation is again that at Planck energy, where tangles are tight. And again, it is easiest to explore the open trefoil that describes the essential aspects of the weak W boson. Any absorption of a Planck energy photon by a W boson is expected to lead to a rotation of the trefoil. We can thus average, over all orientations of the open trefoil, the rotation angle required to add a loop. This average angle should yield the square root of the fine structure constant at Planck energy. The calculation is underway.

The energy dependence of the coupling constants

In nature, coupling constants, like masses and mixing angles, change, i.e., run, with energy. All other physical observables, such as spin, parities or other quantum numbers, are found not to change with energy. For coupling constants between everyday energy and about 100 GeV, the measurement results of the running agree with the prediction from quantum field theory.

The strand model predicts that coupling constants, like masses and mixing angles, change with energy, because they are quantities that depend on the *geometry* of the underlying particle tangles. We also note that the strand model predicts running *only* for these three types of observables; all the other observables – spin, parities or other quantum numbers – are predicted to depend on the *topology* of the particle tangles, and thus to be independent of energy.

In the standard model of particle physics, the running of the electromagnetic and weak coupling constants – the slope in Figure 99 – depends on the number of existing Higgs boson types. The strand model predicts that the number is zero, and thus that measuring the running of the constants can check the number of Higgs bosons. Unfortunately, the difference is small; for the electromagnetic coupling, the slope changes by around 2% if one Higgs exists. But in future, such a measurement accuracy might be possible. Ref. 215

So far, the strand model thus does not contradict observations. We now explore the details. We do this by distinguishing two energy ranges: energies much lower than the Planck energy, and energies near the Planck energy.

Predictions at low energy – comparing coupling constants

At energies much smaller than the Planck energy, such as everyday energies, the strand model implies that the average tangle core size for each particle is of the order of the position uncertainty. In other words, any thickness of the strands – real or effective – can be neglected at low energies.

At low energies, the predictions of the strand model are those that appear when the Planck length is set to zero. In particular, at low energy, the average strand length within a particle tangle core is of the order of the de Broglie wavelength. Everyday energy thus implies *large* and *loose* tangle cores.

At low energies, shape fluctuation can lead to each Reidemeister move. The probabilities of such shape deformations will scale with a power of the average strand length in the tangle core. *Higher* Reidemeister moves will scale with *larger* power values. In particular, the longer the strand – i.e., the lower the energy – the more the relative probability for the higher Reidemeister moves will increase.

In summary, the strand model predicts that at low energy, the strong nuclear interaction, due to the third Reidemeister move, is the strongest gauge interaction, followed by the weak nuclear interaction, due to the second Reidemeister move, in turn followed by the electromagnetic interaction. This prediction matches observations.

The running of the coupling constants

The strand model proposes a new view on the screening and antiscreening effects that are part of quantum field theory. These effects are seen as consequences of the statistics of shape deformations for loose tangle cores. Since these statistical effects can in principle be calculated, it is expected that such calculations can be compared with the predictions of quantum field theory shown in Figure 99. This work is in progress. A few results, however, can be deduced without any computer calculations.

In the strand model, the electromagnetic interaction is due to the first Reidemeister move, the twist. For a charged particle, the average difference in right and left twists determines the effective charge. It is expected that this difference *decreases* when the strand core is loose, because the loose strands will *wash out* the differences due to the chirality of the tangle. In other words, the strand model predicts that the electromagnetic coupling increases with energy, as is observed.

For the nuclear interactions, the washing out effect is not expected; on the contrary, as just explained, the effective nuclear coupling constants are expected to decrease with energy, because the effective number of crossings is expected to decrease.

In other words, the strand model predicts the observed signs for the slopes of the coupling constants in Figure 99.

Predictions at Planck energy

At energies near the Planck energy, strand thickness effects are expected to play a role. At these high energies, to make the strand model self-consistent, strands are expected to have an effective diameter given by the Planck length. Therefore, deviations from the energy dependence predicted by quantum field theory are expected near the Planck energy. Let us try to estimate these deviations.

At such high energies, quantum field theory breaks down. In addition, experiments at such high energies are impossible. We must look for another way to check the model.

We continue with the approximation that 3d writhe determines electric charge. We mentioned above that the 3d writhe of random tangles and the 3d writhe of tight tangles differ only by a very small percentage, of the order of 2%. In a rough approximation, we can set the two values equal to another. *The shape of a* tight *tangle is thus predicted to determine, to within a few percent, the gauge coupling constants at Planck energy.*

The strand model makes thus two important statements about coupling constants. First of all, at Planck energy, the difference between the strand model and quantum field theory are in the percentage range. *Secondly, the stand model predicts that the values of the couplings at Planck energy can be approximated by determining the probability of twists, pokes and slides for* tight *knots.* The last prediction allows the calculation of coupling constants in a simple way; this approach is under investigation.

Open challenge: Calculate coupling constants ab initio

Challenge 161 ny

Page 303

Calculating the coupling constants ab initio, by determining the statistics of strand fluctuations, will allow to check the statements of this section. The effects of the various tangle family members have also to be taken into account, as in the strand model, each particle is described by a family of tangles. It seems, though, that family members have similar effects on charge q and effective coupling $q\sqrt{\alpha}$, so that the family issue can be neglected in first order calculations.

THE FINAL SUMMARY ON THE MILLENNIUM ISSUES

In this chapter, we have deduced that strands predict exactly three generations of leptons and of quarks, precisely those gauge bosons that are known, and precisely those values of their quantum numbers that are observed. We have found that tangles of strands allow to calculate the masses of particles, their coupling constants, their mixing angles and the CP violating phases. First rough estimates of these values agree with the (much more precise) experimental data. Computer calculations will allow to improve these checks in the near future.

These results allow us to sum up our adventure in three statements:

1. *Strands solve all open issues.* With one simple fundamental principle, the strand model solves or at least proposes a way to solve *all* issues from the millennium list of open issues in fundamental physics. All fundamental constants can be calculated with strands. Page 18
2. *Strands agree with all observations.* In particular, the strand model implies that general relativity, quantum theory and the standard model of elementary particles are a *precise* description of motion for all practical purposes.
3. *Nothing new will be discovered in fundamental physics.* Unexpectedly but convincingly, strands predict that general relativity, quantum theory and the standard model of elementary particles are a *complete* description of motion for all practical purposes.

We have not yet literally reached the top of Motion Mountain – because certain calculations are not yet precise enough – but if no cloud has played a trick on us, we have seen the top from nearby. The playful spirit that we invoked at the start has been a good guide. Page 21

EXPERIMENTAL PREDICTIONS OF THE STRAND MODEL

> Es gibt viele Theorien,
> die sich jedem Test entziehen.
> Diese aber kann man checken,
> elend wird sie dann verrecken.*
>
> Anonymous

Many experiments around the world are searching for effects that are unexplained by the standard model of particle physics. All these experiments are testing the strand model. In fact, most people working on these experiments have not heard about the strand model, so that there is not even the danger of unconscious bias.

The strand model predicts that apart from the gauge bosons, the quarks, the leptons and the graviton, *no other* elementary particles exist in nature. In particular, the strand model predicts that of the dozens of particles conjectured in the past – such as Higgs bosons, axions, magnetic monopoles, dyons, superpartners, knotted solitons, other elementary gauge bosons – none exists in nature. For example, the strand model predicts that dark matter is made of ordinary matter and black holes. The most important predictions of the strand model that we deduced in our adventure are listed in Table 14.

---

* No adequate translation possible.

TABLE 14 The main predictions of the strand model. Predictions that are **unique** to the strand model are typeset in **bold** typeface.

	Experiment	Prediction (from 2008/2009)	Status (early 2011)
Page 33	Planck units	are limit values	none has been exceeded, but more checks are possible.
Page 286	Higgs boson	does not exist	not yet found.
Page 323	Running of the coupling constants	implies no Higgs	not yet falsified.
Page 286	Unitarity of longitudinal W and Z boson scattering	is maintained	is observed.
Page 286	**Non-local and non-perturbative effects in W and Z boson scattering**	**will be observed at the Large Hadron Collider**	**not yet tested.**
Page 303	**Unknown fermions (supersymmetric particles, magnetic monopoles, dyons, heavy neutrinos for the see-saw mechanism, etc.)**	**do not exist**	**none found yet.**
Page 274	**Unknown bosons (other gauge bosons, supersymmetric particles, axions, etc.)**	**do not exist**	**none found yet.**
Page 246, page 278	**Unknown interactions and symmetries, grand unification, supersymmetry, quantum groups, technicolour**	**do not exist**	**none found yet.**
Page 273	**Particle masses, mixing angles and coupling constants**	**are calculable by modifying existing software packages**	**not yet tested.**
Page 273	Particle masses, mixing angles and coupling constants	are constant in time	confirmed by experiment.
Page 313	Mixing matrix for quarks	is unitary	is observed.
Page 315	Mixing matrix for neutrinos	is unitary	no data yet.
Page 246	Electric dipole moments of elementary particles	have extremely small, standard model values	still too small to be measured.
Page 295	Neutrinos	are Dirac particles	not yet confirmed.
Page 286	Neutrino-less double beta decay	does not exist	not yet found.
Page 298	Tetraquarks	exist	likely.
Page 283, page 295	Glueballs	probably do not exist	not yet found.
Page 246	Proton decay	occurs at extremely small, standard model rates	not yet found.
Page 296	Neutron decay	follows the standard model	no deviations found.
Page 296	Neutron charge	vanishes	none observed.
Page 246	Neutron-antineutron oscillations	occur at extremely small, standard model rates	not yet found.
Page 292	Hadron form factors	can be calculated ab initio	not yet calculated.
Page 305	**Dark matter**	**is conventional matter plus black holes**	**data is inconclusive.**

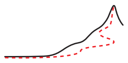
TABLE 14 (Continued) The main predictions of the strand model. Predictions that are **unique** to the strand model are typeset in **bold** typeface.

	Experiment	Prediction (from 2008/2009)	Status (early 2011)
Page 273	**Standard model of particle physics**	**is essentially correct, with deviations for the scattering of longitudinal vector bosons**	**not yet falsified.**
Page 148	Additional dimensions	do not exist	not observed.
Page 148	Non-commutative space-time	does not exist	not observed.
Page 259	General relativity	is correct at all accessible energies	no deviation found.
Page 259	Short-distance deviations from universal gravitation	do not exist	no deviation found.
Page 269	Cosmological constant (dark energy)	is small and positive	is observed.
Page 269	Cosmological constant (dark energy)	decreases with time	data is inconclusive.
Page 268	Cosmological inflation	did not occur	data is inconclusive.
Page 271	Cosmic topology	is trivial	as observed.
Page 258	Space-time singularities, cosmic strings, wormholes, time-like loops, negative energy regions	do not exist	none observed.
Page 262, page 246	Quantum gravity effects	will not be found	not observed yet.
	In summary: all motion	**results from strands**	**not yet falsified.**

The strand model consistently deduces all its experimental predictions from a single, simple fundamental principle: events and Planck units are due to crossing switches of strands. There is no way to change the predictions. The strand model is unmodifiable and simple.

If any experiment ever contradicts the strand model, the model is doomed. At present, the experimental predictions made by the strand model are quite unpopular. Practically all other attempts at unification predict the existence of yet undiscovered particles and effects. However, so far, not a single prediction of the strand model contradicts experiment.

> “Tutto quel che vedete, lo devo agli spaghetti.* ”
> Sofia Loren



---

* ‘Everything you see, I owe it to spaghetti.’

Chapter 13

THE TOP OF MOTION MOUNTAIN

> All things are full of gods.
>
> Thales of Miletus, *c.* 585 BCE

Who am I? Where do I come from? What shall I do? Where does the world come from? Can the whole world really come to an end? What will happen in future? What is beauty? All these questions have a common aspect: they are questions about motion. Our search for answers led us to study motion in all its details. In this quest, every increase in the precision of our description of motion was a step towards the peak of Motion Mountain. Now that we are at the top of the mountain, we can savour what we have achieved and recall the emotions that we have experienced.

In our ascent, we have learned how we move, how we experience our environment, how we grow, what parts we are made of, and how our actions and our convictions about them can be understood. We have learned a lot about the history and a bit about the future of matter, of radiation and of space. We have experienced and understood the many ways in which beauty appears in nature: as colours, as shapes, as rhythms and most of all: as simplicity.

Savouring our achievement means that first of all, we now can look back to where we came from. Then we enjoy the view we are offered and look out for what we could not see before. After that, we search for what is still hidden from our sight. And finally, we take a different path back down to where we live.

OUR PATH TO THE TOP

> The labour we delight in physics pain.
>
> William Shakespeare, *Macbeth*.

Our walk had a simple aim: to talk accurately about all motion. This 2500 year old quest drove us to the top of this mountain. We can summarize our path in three legs: everyday life, general relativity plus quantum theory, and unification.

Everyday life: the rule of infinity

Galilean physics is the description of everyday life. We learned Galilean physics between our birth and secondary school. Galilean physics is the exploration and description of

Ref. 1, Ref. 3 the motion of stones, water, trees, heat, the weather, electricity and light. To achieve this description of our environment, our first and main act in life is to partition experience into experienceS. In other words, our first intellectual act is the invention of *parts*; we invented the *plural*.

The act of partitioning allows us to define sequences among our experiences, and thus to define the concept of *time*. The concept of *space* arises similarly by our possibility to distinguish observations that occur at the same time. By comparing parts with other parts, we define *measurement*. Using all of this, we become able to define *velocity*, *mass* and *electric charge*, among others. These allow us to introduce *action*, the quantity that quantifies change.

For a simple description of observations, we assume that division is possible without end: thus we introduce the infinitely small. We also assume that widening our scope of observation is possible without end. Thus we introduce the infinitely large. Defining parts thus leads us to introduce infinity.

Using parts and, with them, the infinitely small and the infinitely large, we discover that everyday motion has six main properties: it is continuous, conserved, relative, reversible, mirror-invariant and lazy. Motion is lazy because it produces as little change as possible.

Nature minimizes change. This is Galilean physics, or everyday motion, in one statement. It allows us to describe all our everyday experiences with stones, fluids and electric current. The idea of change-minimizing motion implies that nature is continuous and predictable.

RELATIVITY AND QUANTUM THEORY: THE ABSENCE OF INFINITY

> “Vorin haben wir gesehen, daß in der Wirklichkeit das Unendliche nirgends zu finden ist, was für Erfahrungen und Beobachtungen und welcherlei Wissenschaft wir auch heranziehen.*”
>
> David Hilbert

The idea that nature offers an infinite range of possibilities is often voiced with deep personal conviction. However, the results of relativity and quantum theory show the opposite. In nature, speeds, forces, sizes, ages and actions are limited. No quantity in nature is infinitely large or infinitely small. No quantity in nature is defined with infinite precision. There never are infinitely many examples of a situation; the number of possibilities is always finite. The world around us is not infinite; neither its size, nor its age, nor its content. *Nature is not infinite.* This is general relativity and quantum theory in one statement. Ref. 2, Ref. 4

Relativity and quantum theory show that the idea of infinity appears only in *approximate* descriptions of nature; it disappears when talking with precision. Nothing in nature is infinite. For example, we found that the sky is dark at night because nature is not infinite. And we found that quantum theory has probabilities because there is a smallest action value in nature. In fact, the statement that a quantity is infinitely large of infinitely

---

* ‘Above we have seen that in reality, the infinite is nowehere to be found, whatever experiences and observations and whatever knowledge we appeal to.’ Ref. 243

small cannot be confirmed or reproduced by any experiment. Worse, such a statement is falsified by every measurement. In short, we found that infinity is a fantasy of the human mind. In nature, it does not appear. *Infinity is always a lie.*

The number of particles, their possible positions, the states they can have, our brain, our creativity, our possible thoughts: all this is not infinite. Nevertheless, quantum theory and relativity changed the world: they allowed building ultrasound imaging, magnetic resonance imaging, satellite navigation systems, music players and the internet.

Ref. 4

Despite these developments, nothing in our environment is infinite; neither our life, nor our experiences, nor our memories, not even our dreams or our fantasies. Neither the information necessary to describe the universe, nor the paper to write down the formulae, nor the necessary ink, nor the time necessary to understand the formulae is infinite. Nature is not infinite. On the other hand, we also know that the illusion of the existence of infinity is one the most persistent prejudices and myths ever conceived. Why did we use it in the first place?

The habit to use infinity to describe the world has many emotional reasons. For some, it reflects the deep-rooted experience of smallness that we carry within us as a remnant our personal history, when the world seemed so large and powerful. For others, the idea of our smallness allows us to deny somehow the responsibility for our actions or the existence of death. For others again, the idea of a finite universe often, at a first glance, produces deception, disbelief and discouragement. The absence of infinity means that we cannot achieve everything we want, and that our dreams and our possibilities are limited. Clinging to the idea of infinity is a way to avoid confronting this reality.

However, once we face and accept the absence of infinity, we make a powerful experience. We gain in strength. We are freed from the power of those who use this myth to put themselves above others. It is an illuminating experience to reread all those sentences on nature, on the world and on the universe containing the term 'infinite', knowing that they are incorrect, and then clearly experience the manipulations behind them. The desire to make others bow to what is called the infinite is a common type of human violence.

Challenge 162 e

At first, the demise of infinity might also bring panic fear, because it can appear as a lack of guidance. But at closer inspection, the absence of infinity brings strength. Indeed, the elimination of infinity takes from people one of the deepest fears: the fear of being weak and insignificant.

Moreover, once we face the finitude of nature, we react like in all those situations in which we encounter a boundary: the limit becomes a challenge. For example, the experience that all bodies unavoidably fall makes parachuting so thrilling. The recognition that our life is finite produces the fire to live it to the full. The knowledge of death gives meaning to our actions. In an infinite life, every act could be postponed without any consequence. The disappearance of infinity generates creativity. A world without limits is discouraging and depressing. Infinity is empty; finitude is a source of strength and pours passion into our life. Only the finitude of the world ensures that every additional step in life brings us forward. Only in a finite universe is progress possible and sensible. Who is wiser, the one who denies limits, or the one who accepts them? And who lives more intensely?

UNIFICATION: THE ABSENCE OF FINITUDE

> Pray be always in motion. Early in the morning go and see things; and the rest of the day go and see people. If you stay but a week at a place, and that an insignificant one, see, however, all that is to be seen there; know as many people, and get into as many houses as ever you can.
>
> Philip D. Stanhope, *Letters to his Son – on the Fine Art of Becoming a Man of the World and a Gentleman.*

The last part of our trip, described in this text, produced an unexpected result. Not only is nature not infinite; nature is not finite either. None of the quantities which were supposed to be finite turn out to be so. Finitude turns out to be an approximation, or better, an illusion, though a subtle one. *Nature is not finite.* This is the unification of physics in one statement.

Precise observation shows that nothing in nature can be counted. If nature were finite it would have to be (described by) a set. However, the exploration of Planck scales shows that such a description is intrinsically incomplete and inaccurate. Indeed, a description of nature by a set can never explain the number of its elements, and thus cannot explain finitude itself. In other words, any idea that tries to describe nature as finite is a belief, and is never correct. *Finitude is a lie.*

Page 130

We thus lost our security of thought a second time. Nature is not infinite, and nature is not finite. We explored the possibilities left over and found that only one option is left: *Nature is indivisible.* In other words, all parts that we experience are approximations. Both finitude and infinity are approximation of nature. All distinctions are approximate. This central conclusion solved the remaining open issues about motion.*Nature has no parts.*

Recognizing all distinctions as being approximate abolishes the distinction between the permanent aspects of nature ('objects', described by mass, charge, spin, etc.) and the changing aspects ('states', described by position, momentum, energy). Taking all distinctions as approximate introduces extended constituents: fluctuating strands. Looking even closer, these extended constituents are all the same one. Space, formally only used to describe states, also acquires changing aspects: it is made from fluctuating strands. Also properties like mass or charge, which formally were seen as static, become aspects of the ever changing interplay between these fundamental constituents. Describing nature as a fluctuating strand allows us to answer all questions left open by quantum theory and general relativity.

In a sense, the merging of objects and states is a resolution of the contrasting views on motion of the Greek thinkers Parmenides – 'there is no motion', i.e., in physical language, 'there are no states, there is only permanence' – and Heraclitus – 'everything moves', i.e., in physical language 'there is no permanence, there are only states'. Both turn out to be right. We can thus sum up the progress during our adventure of physics in the following way:

Strands unify physics. In particular, strands extend our views on quantum theory and mathematical physics, on particle physics and field theory, on axiomatic physics and algebraic physics, on polymer physics and gauge theory, on general relativity and cosmology.

TABLE 15 The progress of physics.

Step 1	Galilean Physics	Nature is continuous.	We live in Galilean space.
Step 2	Relativity	Nature has no infinitely large.	We live in Riemannian space.
Step 3	Quantum field theory	Nature has no infinitely small.	We live in a Hilbert/Fock space.
Step 4	Unification	Nature is not finite. Nature has no parts.	We do not live in any space; we are space.

NEW SIGHTS

> “Die Natur kann besser Physik als der beste Physiker.*”
>
> Carl Ramsauer

Modelling nature as a complicated web of fluctuating strands allowed us to describe at the same time empty space, matter, radiation, horizons, kefir, stars, children and all our other observations. In short, all everyday experiences are consequence of everything in nature being made of connected strands. Let us explore some of the new sights opened up by this result.

THE BEAUTY OF STRANDS

> “Someday, surely, we will see the principle underlying existence itself as so simple, so beautiful, so obvious, that we will all say to each other, “Oh, how could we all have been so blind, so long.””
>
> John Wheeler, *A journey into gravity and spacetime.*

Describing everything as connected does not come natural to us humans. After all, in our life, we perform only one act: to partition. We define pluralities. There is no way we can avoid doing this. To observe, to think, to talk, to take a decision, to move, to suffer, to love or to enjoy life is impossible without partitioning.

Our walk showed us that there are limits to the ability to distinguish. Any kind of partitioning is always approximate. In fact, most people can summarize their personal experience by saying that they learned to make finer and finer distinctions. However, talking with highest precision about a part of the world inevitably leads to talk about the whole universe. The situation resembles a person who gets a piece of rope in his hand, and by following it, discovers a large net. He continues to pull and finally discovers that everything, including himself, is part of the net.

For the strand model, the term ‘theory of everything’ is therefore not acceptable. Nature cannot be divided into ‘things’. In nature, things are never separable. There is no

---

* ‘Nature knows physics better than the best physicist.’ Carl Ramsauer (1879–1955) was a German physicist and the first person to discover that electrons behave as waves.

way to speak of 'every' thing; there are no sets, no elements and no parts in nature. A theory describing all of nature cannot be one of 'everything', as 'things' are only approximate entities: properly speaking, they do not exist. The strand model is not a theory of everything; it is the *final theory*.

The strand model shows: being in motion is intrinsic to being a part. Parts, being approximate, are always in motion. As soon as we divide, we observe motion. The act of dividing, of partitioning, of defining parts is the very one which produces order out of chaos. Strands force us to rethink this habit.

Despite being so tough to grasp, strands yield a precise description of motion that unifies quantum field theory and general relativity. The strand model for the unification of motion is both simple and powerful. There are no free parameters. There are no questions left. Our view from the top of the mountain is thus complete. No uncertainty, no darkness, no fear and no insecurity are left over.

CAN THE STRAND MODEL BE GENERALIZED?

Page 159 As mentioned above, mathematical physicists are fond of *generalizing* models. Despite this fondness, we required that any final, unified description must be unique: any final, unified description must be impossible to modify or to generalize. In particular, a final theory must neither be a generalization of particle physics nor of general relativity. Let us check this.

What is a requirement to one person, is a criticism to another. A number of researchers deeply dislike the strand model precisely because it doesn't generalize previous theories and because it cannot be generalized. This attitude deserves respect, as it is born from the admiration for the ancient masters of physics. However, the strand model points into a different direction.

The strand model is not a generalization of general relativity: the definitions of curvature, of gravitons and of horizons differ radically from general relativity's approach. The strand model is also not a generalization of particle physics: the definitions of particle and of interactions differ radically from the old concepts of field theory. In short, the strand model doesn't generalize previous theories.

But what about the second requirement for a unified theory? Can the strand model be generalized? We have seen that the model does not work in more spatial dimensions, does not work with more families of quarks, does not work with more interactions, and does not work with other evolution equations in general relativity or particle physics. The strand model does not work with other fundamental constituents, such as bifurcating entities, membranes, bands, or networks. (Though it does work with *funnels*, as explained earlier on, but this description is equivalent to that with strands.) Obviously, exploring all possible variations and modifications remains a challenge for the years to come. If an *inequivalent* modification of the strand model can be found, the strand model instantly loses its value: in that case, it would need to be shelved as a failure. Only a *unique* unified model can be correct.

Page 164

Challenge 163 r

In summary, one of the beautiful aspects of the strand model is its radical departure from twentieth-century physics in its basic concepts, combined with its almost incredible uniqueness. No generalization, no specialization and no modification of the strand model seems possible. In short, the strand model qualifies as a unified, final theory. Ref. 157

What is nature?

> “Nature is what is whole in each of its parts.”
> Hermes Trismegistos, *Book of Twenty-four Philosophers*.

Ref. 244

At the end of our long adventure, we discovered that nature is not a set: everything is connected. Nature is only *approximately* a set. The universe has no topology, because space-time is not a manifold. Nevertheless, the approximate topology of the universe is that of an open Riemannian space. The universe has no definite particle number, because the universe is not a container; the universe is made of the same stuff of which particles are made. Nevertheless, the approximate particle density in the universe can be deduced.

We thus arrive at the (slightly edited) summary given around the year 1200 by the author that wrote under the pen name Hermes Trismegistos: *Nature is what is whole in each of its parts.* But in contrast to that author, we now also know how to draw testable conclusions from the statement.

Quantum theory and the nature of matter

> “In everything there is something of everything.”
> Anaxagoras of Clazimenes (500–428 BCE)

The strand model shows that as soon as we separate the universe into space-time and the rest, i.e., as soon as we introduce the coordinates x and t, quantum mechanics appears automatically. More precisely, *quantum effects are effects of extension.* Quantum theory appears when we realize that observations are composed of smallest events due to crossing switches, each with a change given by the quantum of action. All events and observations appear through the fluctuations of the strand that composes nature.

We found that *matter is made of tangled strands*. In fact, the correct way would be to say: matter is made of tangled strand *segments*. This connection leads to Schrödinger's equation and to Dirac's equation.

Insofar as matter is of the same fabric as the vacuum, we can rightly say: *matter is made of nothing*. But the most appropriate answer arises when we realize that matter is not made from something, but that matter is a certain aspect of the *whole* of nature. Unification showed that every single elementary particle results from an arrangement which involves the whole of nature, or, if we prefer, the entire universe. In other words, we can equally say: *matter is made of everything.*

Cosmology

The strand model also showed us how to deduce general relativity. The strand model clarified the fabric of horizons and explained the three dimensions of space. Most fascinating is the idea of a universe as the product of a single strand. A single strand implies that there was nothing before the big bang, and that there is nothing outside the night sky: no 'multiverse' and no hidden worlds of any kind. And the fluctuating strand explains all observations of the universe at large.

The 'big bang' is the name for what we observe if we try to make observations approaching the limits of nature. It appears automatically from the strand model whenever we observe nature at the most distant times, the largest distances or at the largest energies: 'big bang' is the name for Planck scale physics.

The universe consists of a single strand. There are many particles in nature, because the strand is tangled up in complicated ways. What we call the 'horizon' of the universe is the place where new tangles appear.

The belief that the big bang or the horizon are examples of creation is incorrect. What happened at the big bang still happens at the horizon. Both the big bang and the black sky at night are nature's way to tell us: 'Galilean physics is approximate! Quantum theory is approximate! General relativity is approximate!'

WHY IS THERE ANYTHING INSTEAD OF NOTHING?

> “Was man nicht träumen kann, hat keine Wirklichkeit.*”
>
> Ernst Erich Nossak

Asking why there is anything instead of nothing leads to a inescapable answer. We can start from the definitions of the terms used in our walk: we know that 'is' means 'able to interact', and 'anything' means a 'part in relation to others'. We then get a first answer: there are things, because that is the way we defined them: things *are*. If we look further and ask how we deduced these definitions, we can get a second answer: things *are* only because we also *are* (things). But this is idle talk, which moreover depends on the exact meaning one gives to the terms of the question. Our walk gave us a fresh and final answer: there is *no* difference between anything and nothing. There is also no difference between 'being' and 'not being' in nature. Page 54

In short, the question of the section title does not pose an alternative. This conclusion might be the one which instils us with the largest possible amount of awe. It also shows most clearly how limited our human imagination can be.

MUSINGS ABOUT UNIFICATION AND STRANDS

> “Continuing motion masters coldness.
> Continuing rest masters heat.
> Motion based on rest:
> Measure of the all-happening for the single one.”
>
> Lao Tse, *Tao Te King*, XXXXV.

All is made from one sort of thing: all is one substance. This idea, *monism*, sounds a lot like what the Dutch philosopher Baruch Spinoza (1632–1677) held as conviction. Monism, though mixed up with the idea of god, is also the basis of the philosophical ideas that Gottfried Wilhelm Leibniz (1646–1716) presents in his text *La Monadologie*.

* *

Any complete theory of motion, also the strand model, is built on a single statement about nature: The *many* exists only approximately. Nature is approximately multiple. The etymological meaning of the term 'multiple' is 'it has many folds'; in a very specific sense, nature thus has many folds. Ref. 245

---

* 'What cannot be dreamed, has no reality.'

* *

Any precise description of nature is free of arbitrary choices, because the divisions that we have to make in order to think are all common to everybody, and logically inescapable. Since physics is a consequence of this division, it is also 'theory-free' and 'interpretation-free'. This consequence of the final theory will drive most philosophers up the wall.

* *

For over a century, physics students have been bombarded with the statement: 'Symmetries are beautiful'. Every expert on beauty, be it a painter, an architect, a sculptor, a musician, a photographer or a designer, fully and completely disagrees, and rightly so. Beauty has no relation to symmetry. Whoever says the contrary is blocking out his experiences of a beautiful landscape, of a beautiful human figure or of a beautiful work of art.

The correct statement is: 'Symmetries simplify descriptions.' Symmetries simplify physical theories. In particular, the search for simplicity, not the search for beauty, has always driven the progress of fundamental theoretical physics.

* *

The description of nature with strands is surprisingly simple, since it uses so few basic concepts. Is this result astonishing? In our daily life, we describe our experiences with the help of a few thousand words, e.g. taking them from the roughly 350 000 words which make up the English language, or from a similar number from another language. This set is sufficient to talk about everything, from love to suffering, from beauty to happiness. And these terms are constructed from no more than about 35 basic ones, as we have seen already. We should not be too surprised that we can in fact talk about the whole universe using only a few basic concepts: the act and the results of distinction, or more specifically, a basic event – the crossing switch – and its observation.

Vol. III, page 193

* *

Almost all discoveries in physics were made at least 30 years too late. The same is true for the strand model. If we compare the strand model with what many physicists believed in the twentieth century, we can see why: researchers had too many wrong ideas about unification. All these wrong ideas can be summarized in the following statement:

Page 21

— 'Unification requires generalization of existing theories.'

This statement is subtle: it was never expressed explicitely but widely believed. But it is wrong, and it led many astray. On the other hand, the development of the strand model also followed a specific guiding idea, namely:

— 'Unification requires simplification.'

Hopefully this guiding idea will not become a dogma itself; in many domains of life, simplification can do a lot of harm.

* *

The strand model shows that achieving unification is not a feat requiring difficult abstraction. Unification was not hidden in some almost inaccessible place that can reached only by a few select, well-trained research scientists. No, unification is accessible to everyone

FIGURE 100 Motion Mountain does not resemble Cerro Torre, but a gentle hill (© Davide Brighenti, Myriam70)

who has a basic knowledge of nature and of physics. No Ph.D. in theoretical physics is needed to understand or to enjoy it. The knowledge presented in the volumes of this series is sufficient.

When Andrew Wiles first proved Fermat's last theorem after three centuries of attempts, he explained that his search for a proof was like the exploration of a dark mansion. And seen the difficulties he had to overcome, the analogy was fitting. Recalling how many more people have already searched for unification without success, the first reaction is to compare the search for unification to the exploration of something even bigger, such as a complex dark cave system. But that analogy was only partially helpful. In contrast to the proof of Fermat's theorem, the goal of the quest for unification turned out to be lying out in the open. Most researchers simply overlooked it, because they were convinced that the goal was carefully hidden, in the dark, and hard to reach. It was not.

The adventure of climbing Motion Mountain is thus not comparable to climbing Cerro Torre, which might be the toughest and most spectacular challenge that nature offers to mountain climbers. Figure 100 gives an impression of the peak. Neither does Motion Mountain resemble the peak from the Himalaya shown on the cover. Climbing Motion Mountain is more like walking up a gentle green hill, alone, with a serene mind, on a sunny day, while enjoying the surrounding beauty of nature.

* *

Page 78

The strand model also settles all questions about *determinism*. Quantum theory and general relativity are deterministic. Nevertheless, when both descriptions are combined, time turns out to be an approximate, low-energy concept. The same applies to determinism. Even though nature is deterministic for all practical purposes, determinism shares the fate of all its conceivable opposites, such as fundamental randomness, indeterminism of all kinds, existence of wonders, creation out of nothing, or divine intervention: determinism, like all its alternatives, is an *incorrect* description of nature at the Planck scale.

THE ELIMINATION OF INDUCTION

> “Cum iam profeceris tantum, ut sit tibi etiam tui reverentia, licebit dimittas pedagogum.*”
>
> Seneca

The theory of motion has a consequence worth mentioning in detail: its lack of infinity and its lack of finitude eliminate the necessity of induction. This conclusion is of importance for general discussions on man’s grasp of nature.

In physics, as in the other natural sciences, there is a tradition to state that a certain ‘law’ of nature is valid in *all* cases. In these statements, ‘all’ means ‘for all values of the quantities appearing’. As a concrete example, the ‘law’ of universal gravitation is always claimed to be the same here and today, as well as at *all* other places and times, such as on the other end of the universe and in a few thousand years. The full list of such all-claims is part of the millennium list of open issues in twentieth-century physics. Page 158

For many decades, the habit of claiming general validity from a limited and finite number of experiences, also called *induction*, has been seen, and rightly so, as a logically dubious manoeuvre, tolerated only because it works. But the developments described in this text show that this method is indeed justified.

First of all, a claim of generality is not that enormous as it may seem, since the number of events that can be distinguished is finite, not infinite. The preceding sections showed that the maximal number N of events that can be distinguished in the universe is of the order of $N = (T_0/t_{\mathrm{Pl}})^4 = 10^{244\pm2}$, T_0 being the age of the universe and t_{Pl} the Planck time. This is a big, but certainly finite number.

The unified description of nature has thus first reduced the various all-claims from an apparently infinite to a finite number of cases, though still involving astronomically large numbers. This change results from the recognition that infinities do not appear in the description of nature. We now know that when talking about nature, ‘all cases’ never means an infinite number.

A second, important result is achieved by the description of nature with strands. In any all-claim, the checking of each of the large number of possibilities is not necessary, since all events result from a single entity, in which we introduce distinctions with our senses and our brain. And the distinctions we introduce imply automatically that the symmetries of nature – the ‘all-claims’ or ‘inductions’ – that are used in the description of motion are correct. Nature does not contain separate parts. Therefore, there is no way that separate parts can behave differently. Induction is a result of the unity of nature.

Ultimately, the possibility to *verify* statements of nature is due to the fact that all the aspects of our experience are *related*. Complete separation is impossible in nature. The

---

* ‘When you have profited so much that you respect yourself you may let go your tutor.’ Seneca writes this is his *Epistulae morales ad Lucilium*, XXV, 6.

verification of all-claims is possible because the strand model achieves the full description of how all 'parts' of nature are related.

The strand model shows that we can talk and think about nature because we are a part of it. The strand model also shows that induction works because everything in nature is related to everything else: nature is one.

WHAT IS STILL HIDDEN?

> That which eludes curiosity can be grasped in action.
>
> Traditional saying.

Where do we come from? Where does the world come from? What will future bring? What is death? All these questions are questions about motion – and its meaning. And like all mountain climbers, we also have to ask: why are we climbing? Like all mountain climbers, we have to admit that climbing, like every other passion, is also a symbolic activity. Climbing is a search for our mother, for meaning, and for ourselves.

To all such questions, the strand model provides only abstract answers: We are a collection of tangled strands. We are everything and nothing. The strand(s) we are made of will continue to fluctuate. Birth, life and death are part of nature. The world is a folded strand that grows in complexity.

Obviously, such abstract answers do not help. Indeed, to achieve a precise description of motion, we essentially studied only the details of moving particles and of bending space. Studying them was a sequence of riddles; but solving these riddles does not provide meaning, not even at the top of Motion Mountain. From the top we cannot see the evolution of complicated systems; in particular, we cannot see or describe the evolution of life, the biological evolution of species, or the growth of a human beings. From the top we cannot see the details down in the valleys of human relations or experiences. In short, strands do not provide meaning. To find meaning, we have to descend back down to real life. Remaining too long on the top of Motion Mountain is not useful.

Vol. I, page 16

A RETURN PATH: JE RÊVE, DONC JE SUIS

> I hate reality. But it is the only place where one can get a good steak.
>
> Woody Allen

Enjoying life and giving meaning to one's life, requires to descend from the top of Motion Mountain. The return path can take various different directions. From a mountain, the most beautiful and direct descent might be the use of a paraglider. After our adventure, we take an equally beautiful way: we leave reality.

The usual trail to study motion, also the one of this text, starts from our ability to talk about nature to somebody else. From this ability we deduced our description of nature, starting from Galilean physics up to the strand model. The same results can be found by requiring to be able to talk about nature to ourselves. Talking to oneself is an example of thinking. We should therefore be able to derive all physics from René Descartes' sentence

Ref. 246 'je pense, donc je suis' – which he translated into Latin as 'cogito ergo sum'. Descartes stressed that this is the only statement of which he is completely sure, in opposition to his observations, of which he is not. He had collected numerous examples in which the senses provide unreliable information.

However, when talking to ourselves, we can make more mistakes than when asking for checks from others. Let us approach this issue in a radically different way. We directly proceed to that situation in which the highest freedom is available and the largest number of mistakes are possible: the world of dreams. If nature would only be a dream, could we deduce from it the complete set of physical knowledge? Let us explore the issue.

- Dreaming implies the use of distinctions, of memory and of sight. Dreams contain *parts* and *motion*.
- Independently on whether dreams are due to previous observations or to fantasies, through memory we can define a sequence among them. The order relation is called *time*. The dream aspects being ordered are called *events*. The set of all (dream) events forms the (dream) *world*. Ref. 247
- In a dream we can have several independent experiences at the same time, e.g. about thirst and about hunger. Sequences thus do not provide a complete classification of experiences. We call the additional distinction *space*. Dream space has three dimensions.* Dreaming thus means to use space and time.
- We can distinguish between dream contents. Distinguishing means that we can count items in dreams. Counting means that we have a way to define measurements. Dreams are thus characterized by something which we can call 'observables'. Dreams are characterized by a *state*.
- Since we can describe dreams, the dream contents exist independently of dream time. We can also imagine the same dream contents at different places and different times in the dream space. There is thus an invariance of dream concepts in space and time. There are thus symmetries in dream space.
- Dream contents can interact. Dreams appear to vary without end. Dreams seem to be infinite.

In other words, a large part of the world of dreams is described by a modified form of *Galilean physics*. We note that the biggest difference between dreams and nature is the lack of conservation. In dreams, observations can appear, disappear, start and stop. We also note that instead of dreams, we could equally explore *films*. Films, like dreams, are described by a modified form of Galilean physics. And films, like dreams, do not follow conservation laws. But dreams teach us much more.

- Dreams show that space can warp.
- Dream motion, as you may want to check, shows a maximum speed. Challenge 164 ny
- Dreams show a strange limit in distance. There is a boundary to our field of vision, even though we do not see it.

Pondering these issues shows that there are *limits* to dreams. In summary, the world of dreams has a maximum size, a maximum speed and three dimensions that can warp. The

---

* Though a few mathematicians state that they can *think* in more than three spatial dimensions, all of them *dream* in three dimensions.

world of dreams and of films is described by a simple form of *general relativity*.

- Both the number of items we can dream of at the same time as well as the memory of previous dreams is finite.
- There are pixels in dreams, though we do not experience them directly. But the existence of a highest number of things we can dream of at the same time means that dream space has a smallest scale.

In summary, the world of dreams has something similar to a minimum change. The world of dreams and that of films is described by a simple form of *quantum theory*. The difference with nature is that in dreams and films, space is discrete from the outset. But there is still more to say about dreams.

- There is no way to say that dream images are made of mathematical points, as there is nothing smaller than pixels.
- In dreams, we cannot clearly distinguish content ('matter') and environment ('space').
- In dreams, fluctuations appear both for images as well as for the background.
- In dreams, sharp distinctions are impossible. Dream space-time cannot be a set.
- Dream motion appears when approximate constancy (over time) is observed.
- In dreams, dimensionality is not clear; for example, two and three dimensions are mixed up.

In summary, the world of dreams seems to behave as if it is described by extended constituents.

We thus conclude this short exploration of the physics of dreams with a fascinating *conjecture*: even if nature would be a dream, an illusion or a fantasy, we might still get most of the results that we discovered in our ascent of Motion Mountain. (What differences with modern physics would be left?) Speaking with tongue in cheek, the fear of our own faults of judgement, so rightly underlined by Descartes and many others after him, might not apply to fundamental physics.

Challenge 165 s

WHAT IS MOTION?

> Deep rest is motion in itself. Its motion rests in itself.
>
> Lao Tse, *Tao Te King*, VI.

We can now answer the question that drove us through our adventure. *Motion* is the observation of crossing switches of the one, unobservable, tangled and fluctuating strand that describes nature.

The observation of crossing switches, the fluctuating strand segments and their (approximate) embedding in a background space result and are possible because also we are made of the same strand that makes up nature. Like all observers, we are limited: like all observers, we are built of fluctuating strand segments. These strands form the elementary particles inside us. These strands and particles lead us to introduce background space. Motion thus appears automatically when approximate parts of nature, such as humans, animals or machines, describe other approximate parts of nature, such as other bodies.

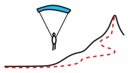
Motion is a consequence of the fact that distinguishing between parts, between bodies, or between bodies and space, is only *approximately* possible. Motion appears as soon as nature is divided up. More clearly, motion appears as soon as we divide the world into parts and then follow them. But this is not a conscious act; our human nature – our senses and our brain – *forces* us to perform this division. As a result, whenever we experience or talk about parts of the universe, we find motion. In other words, *we experience motion because we are limited*: our senses and our brain are made to distinguish. We can't do otherwise. We need this ability for survival and for enjoying life. In a sense, we can say that motion appears as a logical consequence of our limitations; the main limitation is the one that makes us introduce multitudes: elements and sets.

In short, our introduction of the *plural* leads to the observation of motion. Motion is thus the result of our use of (approximate) *parts* to attempt to describe the *unity* of nature. Since the observation of motion results from approximations, motion is an 'artefact' of local interactions. Thus, in a certain sense, motion is an illusion. We seem to confirm what Zeno of Elea stated 2500 years ago. But in contrast to Zeno's pessimistic view, we now have a fascinating spectrum of results and tools at our disposition. They allow us to describe our environment with high precision. Most of all, these tools allow us to improve it.

Vol. I, page 15

Ref. 248

> “All the great things that have happened in the world first took place in a person's imagination, and how tomorrow's world will look like will largely depend on the power of imagination of those who are just learning to read right now.”
> Astrid Lindgren



POSTFACE

Perhaps once you will read Plato’s *Phaedrus*, one of the beautiful philosophical Greek texts. In it, Socrates is made to say that he almost never left the city walls because to him, as a ‘lover of learning, trees and the open country do not teach anything, whereas men in the town do.’ This is a veiled critique of Democritus, the most important and famous philosopher in Greece during Plato’s time. Democritus was the natural philosopher par excellence, and arguably had learned from nature – with its trees and open country – more than anybody else after him.

After this mountain ascent you can decide for yourself which of these two approaches is more congenial to you. It might be useful to know that Aristotle refused to choose and cultivated them both. There is no alternative in life to following one’s own mind, and to enjoy doing so. If you enjoyed this particular trip, show it to your friends. For yourself, after this walk, sense intensively the pleasure of having accomplished something important. Many before you did not have the occasion. Enjoy the beauty of the view offered. Enjoy the vastness of horizon it provides. Enjoy the impressions that it creates inside you. Collect them and rest. You will have a treasure that will be useful in many occasions. Then, when you feel the desire of going further, get ready for another of the adventures life has to offer.

* * *

Plato’s *Phaedrus*, written around 380 BCE, is available in many pocket editions. Do not waste your time learning ancient Greek to read it; the translated versions are as beautiful as the original. Half the text is about love and gave rise to the expression ‘platonic love’, even though its original meaning has been strongly distorted in the meantime, as you will find out.

Plato’s lifelong avoidance of the natural sciences had two reasons. First of all, he was jealous of Democritus. Plato never even cites Democritus in his texts. Democritus was the most prolific, daring, admired and successful philosopher of his time (and maybe of all times). Democritus was a keen student of nature. His written works did not survive, because his studies were not congenial to the followers of christianity, and thus they were not copied by the monks in the Middle Ages. The loss of these texts is related to the second reason that kept Plato away from the natural sciences: he wanted to save his life. Plato had learned one thing from men in the town: talking about nature is dangerous. Starting around his lifetime, for over 2000 years people practising the natural sciences were regularly condemned to exile or to death for impiety. Fortunately, this is only rarely the case today. But such violence still occurs, and is worth remembering the dangers that those preceding us had to overcome in order to allow us enjoying this adventure.

Appendix A

KNOT GEOMETRY

The following table provides a terse summary of the mathematics of knot shapes.

TABLE 16 Important properties of knot, links and tangles.

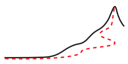
Concept	Defining property	Other properties
Knot / link / tangle	one closed / several closed / one or several open curves, all in 3d without intersections	ropelength is integral of arclength; ropelength is shape-dependent.
Ideal knot, link, tangle (shapes)	tightest possible knot, link or tangle (shapes) assuming a rope of constant diameter that is infinitely flexible and infinitely slippery	at present, all non-trivial ideal shapes are only known approximately; most ideal knots (almost surely) have kinks.
Ribbon or *framing*	short perpendicular (or non-tangent) vector attached at each point of a curve	
Curvature of a curve	inverse curvature radius of ‘touching’ circle	measures departure from straightness, i.e., local bending of a curve.
Normal vector or curvature vector	local vector normal to the curve, in direction of the centre of the ‘touching’ circle, with length given by the curvature	is given by the second and first derivatives of the curve.
Binormal vector	local unit vector normal to the tangent and to the normal/curvature vector	
Torsion	local speed of rotation of the binormal vector; positive (negative) for right-handed (left-handed) helix	measures departure from flatness, i.e., local twisting or local handedness of a curve; essentially a third derivative of the curve.
Frenet frame at a curve point	‘natural’ local orthogonal frame of reference defined by *unit* tangent, *unit* normal/curvature and binormal vector	the Frenet frame differs at each curve point, the Frenet frame is *not* uniquely defined if the curve is locally straight.

TABLE 16 (Continued) Important properties of knot, links and tangles.

CONCEPT	DEFINING PROPERTY	OTHER PROPERTIES
‘Natural’ framing or *Frenet ribbon*	defined by the local normal, i.e., local curvature vector	for a closed curve, it is always closed and two-sided, and thus never a Moebius band.
Linking number between two closed curves	sloppily, number of times that two curves wind around each other, or, equivalently, half the number of times that the curves ‘swap’ position	topological invariant, i.e., shape-independent; $\mathrm{Lk}(K1, K2) = \frac{1}{4\pi} \oint_{K2} \oint_{K1} \frac{\boldsymbol{r}_{12}\,(\mathrm{d}\boldsymbol{r}_1 \times \mathrm{d}\boldsymbol{r}_2)}{\boldsymbol{r}_{12}^3}$.
Linking number for a closed two-sided ribbon	number of times that the edges wind around each other	topologicalinvariant, i.e., shape-independent; always an integer.
Self-linking number or ‘natural’ linking number for a knot	number of times that the edges of the natural/Frenet ribbon wind around each other	not a topological invariant, because of existence of inflection points
Link integral for an open curve	generalization of the linking number for knots to open curves	usually not an an integer.
Twist of a ribbon, open or closed	total angle (in units of 2π) by which the ribbon rotates around the central axis of the ribbon; sloppily said, measures the *local* helicity	vanishes for ribbons that are everywhere flat.
Twist of a curve or knot	total angle (in units of 2π) by which the Frenet frame rotates around the tangent direction, or equivalently, (total) twist of the Frenet ribbon, also called the *total torsion* of the curve	not an integer even in case of knots; depends on curve/knot shape; is different from zero for chiral curves/knots; is zero for achiral curves/knots that have a rigid reflective symmetry; twist and torsion are only equal if the twist is defined with the Frenet ribbon – with other framings they differ; this type of twist has no relation to the first Reidemeister move.
Signed crossing number	sum of positive minus sum of negative crossings in a given oriented 2d projection of a curve or knot (sometimes called ‘2d-writhe’)	always an integer; depends on shape.
2d-writhe of a knot, or *topological writhe*, or *Tait number*	signed crossing number for a *minimal* crossing number diagram/projection (sometimes the term ‘2d-writhe’ is used for the signed crossing number of *any* configuration)	is shape-invariant; is always an integer; differs from 0 for all chiral knots; has the value 3 for the trefoil, 0 for the figure-eight knot, 5 for the 5_1 and 5_2 knots, 2 for the 6_1 knot, 7 for the 7_1 and 7_2 knots, 4 for the 8_1 knot, and 9 for the 9_2 knot.

TABLE 16 (Continued) Important properties of knot, links and tangles.

Concept	Defining property	Other properties
Writhing number or 3d-*writhe* of a knot	average, over all projection directions, of the signed crossing number; sloppily said, measures how wrapped, coiled and chiral a knot is, i.e., measures its *global* helicity	depends on knot shape; usually is not an integer; is different from zero for chiral knots; is zero for achiral knots that have a rigid reflective symmetry; $\mathrm{Wr}(K) = \frac{1}{4\pi} \oint_K \oint_K \frac{\mathbf{r}_{12}\,(\mathrm{d}\mathbf{r}_1 \times \mathrm{d}\mathbf{r}_2)}{r_{12}^3}$; uses no ribbon and thus is independent of the ribbon shape attached to the knot.
Writhe of ideal, alternating knots and of odd-component links	the value is quasi-quantized for alternating knots with small crossing numbers (< 11) in values that differ from $m4/7$ by only a few per cent	is additive under knot addition for knots with small crossing numbers (< 11) within less than 1%.
Writhe of ideal, alternating even-component links	the value is quasi-quantized for alternating links with small crossing numbers (< 11) in values that differ from $2/7 + m4/7$ by only a few percent	
Writhe of a ribbon	sloppily said, measures how wrapped, coiled and chiral a ribbon is, i.e., measures its *global* helicity	
Writhe of an open curve		vanishes for plane curves
Calugareanu's theorem	for any knot K and any ribbon G attached to it, $\mathrm{Lk}(K, G) = \mathrm{Tw}(K, G) + \mathrm{Wr}(K)$	to apply the theorem to *open* curves, a (standardized) closing of curves is required.



CHALLENGE HINTS AND SOLUTIONS

Challenge 1, page 9: Do not hesitate to be demanding and strict. The next edition of the text will benefit from it.

Challenge 2, page 26: Take $\Delta f\ \Delta t \geqslant 1$ and substitute $\Delta l = c/\Delta f$ and $\Delta a = c/\Delta t$.

Challenge 18, page 45: To my knowledge, no such limits have been published. Do it!

Challenge 19, page 45: The system limits cannot be chosen in other ways; after the limits have been corrected, the limits given here should still apply.

Challenge 22, page 46: Just insert numbers to check this.

Challenge 28, page 57: The other domain is the horizon of the universe, as we will see shortly. However, it might be better to argue that there is no other domain, as Planck scales and horizons are not really different, but related situations.

Challenge 29, page 60: Sloppily speaking, such a clock is not able to move its hands in a way that guarantees precise time reading.

Challenge 33, page 76: The final energy E produced by a proton accelerator increases with its radius R roughly as $E \sim R^{1.2}$; as an example, CERN's SPS achieves about 450 GeV for a radius of 740 m. Thus we would get a radius of more than 100 000 light years (larger than our galaxy) for a Planck energy accelerator. Building an accelerator achieving Planck energy is impossible.

Nature has no accelerator of this power, but gets near it. The maximum measured value of cosmic rays, 10^{22} eV, is about one millionth of the Planck energy. The mechanism of acceleration is still obscure. Neither black holes nor the cosmic horizon seem to be sources, for some yet unclear reasons. This issue is still a topic of research.

Challenge 34, page 76: The Planck energy is $E_{\mathrm{Pl}} = \sqrt{\hbar c^5/G} = 2.0\,\mathrm{GJ}$. Car fuel delivers about 43 MJ/kg. Thus the Planck energy corresponds to the energy of 47 kg of car fuel, about a tankful.

Challenge 35, page 77: Not really, as the mass error is equal to the mass only in the Planck case.

Challenge 36, page 77: It is improbable that such deviations can be found, as they are masked by the appearance of quantum gravity effects. However, if you do think that you have a prediction for a deviation, publish it. Page 248

Challenge 38, page 77: There is no gravitation at those energies and there are no particles. There is thus no paradox.

Challenge 39, page 78: The issue is still being debated; a good candidate for a minimum momentum of a single particle is given by $\hbar/R$, where R is the radius of the universe. Is this answer satisfying?

Challenge 40, page 78: The Planck acceleration is given by $a_{\mathrm{Pl}} = \sqrt{c^7/\hbar G} = 5.6 \cdot 10^{51}\ \mathrm{m/s^2}$.

Challenge 41, page 79: All mentioned options could be valid at the same time. The issue is not closed and clear thinking about it is not easy.

Challenge 42, page 79: The precise energy scale is not clear. The scale is either the Planck energy or within a few orders of magnitude from it; the lowest possible energy is thus around a thousandth of the Planck energy.

Challenge 44, page 81: If you can think of an experiment, publish the proposal!

Challenge 45, page 83: Good – then publish it!

Vol. I, page 209

Challenge 46, page 85: The table of aggregates shows this clearly..

Challenge 47, page 86: The cosmic background radiation is a clock in the widest sense of the term.

Challenge 48, page 87: This is told in detail in the section starting on page 33.

Challenge 70, page 101: For the description of nature this is a contradiction. Nevertheless, the term 'universe', 'set of all sets' and other mathematical terms, as well as many religious concepts are of this type.

Challenge 72, page 102: The physical concepts most related to 'monad' are 'strand' and 'universe', as shown in the second half of this text.

Challenge 73, page 102: The macroscopic content of the universe may be observer-dependent. But to speak about many universes is nonsense.

Vol. II, page 231

Challenge 74, page 102: True. Since particles and space are indistinguishable, removing particles means to remove everything.

Vol. III, page 232

Challenge 75, page 102: True. Existence is the ability to interact. If the ability disappears, existence disappears. In other words, 'existence' is a low-energy concept.

Challenge 76, page 104: Plotinus in the *Enneads* has defined 'god' in exactly this way. Later, Augustine in *De Trinitate* and in several other texts, and many subsequent theologians have taken up this view. (See also Thomas Aquinas, *Summa contra gentiles*, 1, 30.) The idea they propose is simple: it is possible to clearly say what 'god' is *not*, but it is impossible to say what 'god' *is*. This statement is also part of the official *Roman Catholic Catechism*: see part one, section one, chapter one, IV, 43, found at www.vatican.va/archive/ENG0015/__PC.HTM. Similar statements are found in Judaism, Hinduism and Buddhism.

The properties common to 'universe' and to 'god' suggest the conclusion that both are the same. Indeed, the analogy between the two concepts can be expanded to a proof. (This exercise is left to the reader.) In fact, this might be the most interesting of all proofs of the existence of 'god', as it lacks all the problems that the more common 'proofs' have. Despite its interest, this proof of equivalence is not found in any book on the topic yet. The reason is twofold. First, the results of modern physics showing that the concept of universe has all these strange properties are not common knowledge yet. Secondly, the result of the proof, the identity of 'god' and the universe – also called *pantheism* – is a heresy for most religions. It thus is an irony that the catholic catechism, together with modern physics, can be used to show that pantheism is a hidden aspect of christianity.

If one is ready to explore the identity of universe and 'god', one finds that a statement like 'god created the universe' translates as 'the universe implies the universe'. The original statement is thus not a lie any more, but is promoted to a tautology. Similar changes appear for many other – but not all – statements using the term 'god'. Enjoy the exploration.

Challenge 77, page 105: If you find one, publish it! And send it to me as well. The previous challenge shows why this issue is interesting.

Challenge 79, page 106: If you find one, publish it and send it also to me.

Challenge 81, page 107: In fact, no length below the Planck length itself plays any role in nature.

Challenge 84, page 110: Any change in rotation speed of the Earth would change the sea level.

Challenge 85, page 111: Just measure the maximum water surface the oil drop can cover, by looking at the surface under a small angle.

Challenge 86, page 112: Keep the fingers less than 1 cm from your eye.

Challenge 87, page 117: As vacuum and matter cannot be distinguished, both share the same properties. In particular, both scatter strongly at high energies.

Challenge 89, page 126: The number of spatial dimensions must be given first, in order to talk about spheres.

Challenge 90, page 130: This is a challenge to you to find out and publish; it is fun, it may bring success, and it would yield an independent check of the results of the section.

Challenge 93, page 138: The lid of a box must obey the indeterminacy relation. It cannot be at perfect rest with respect to the rest of the box.

Challenge 97, page 139: No. Time is continuous only if *either* quantum theory and point particles *or* general relativity and point masses are assumed. The argument shows that only the combination of *both* theories with continuity is impossible.

Challenge 98, page 139: Yes, as nature's inherent measurement errors cannot clearly distinguish between them.

Challenge 99, page 139: Of course.

Challenge 100, page 139: We still have the chance to find the best approximate concepts possible. There is no reason to give up.

Challenge 101, page 139: Here are a few thoughts. A beginning of the big bang does not exist; something similar is given by that piece of continuous entity which is encountered when going backwards in time as much as possible. This has several implications.

- Going backwards in time as far as possible – towards the 'beginning' of time – is the same as zooming to smallest distances: we find a single strand of the amoeba.
- In other words, we speculate that the whole world is one single piece, knotted, branched and fluctuating.
- Going far away into space – to the border of the universe – is like taking a snapshot with a short shutter time: strands everywhere.
- Whenever we sloppily say that extended entities are 'infinite' in size, we only mean that they reach the horizon of the universe.

In summary, no starting point of the big bang exists, because time does not exist there. For the same reason, no initial conditions for particles or space-time exist. In addition, this shows there was no creation involved, since without time and without possibility of choice, the term 'creation' makes no sense.

Challenge 102, page 139: The equivalence follows from the fact that all these processes require Planck energy, Planck measurement precision, Planck curvature, and Planck shutter time.

Challenge 110, page 157: Yes; the appearance of a crossing does not depend on distance or on the number of strands in between.

Challenge 111, page 157: No; more than three dimensions do not allow us to define a crossing switch.

Challenge 112, page 157: If so, let me know. If the generalization is genuine, the strand model is not correct.

Challenge 125, page 204: Yes, as can easily be checked by rereading the definitions with the spinor tangle description in mind.

Challenge 127, page 204: If the strand interpenetration is allowed *generally*, quantum theory is impossible to derive, as the spinor behaviour would not be possible. If strand interpenetration were allowed only *under certain conditions* (such as only for a strand with itself, but not among two different strands), quantum theory might still possible. A similar process lies at the basis of mass generation, as shown in the section on the weak interaction.

Challenge 128, page 205: The belt trick would imply that a wheel rolls over its own blood supply at every second rotation.

Challenge 131, page 221: Julian Schwinger, the discoverer of this formula, has it on his tombstone. A simple derivation, possibly using the approach used in *Buffon's needle* problem, would be a beautiful result.

Challenge 160, page 320: If you plan such a calculation, I would be delighted to help.

Challenge 163, page 334: There is a good chance, however, that such alternatives can be eliminated very quickly. If you cannot do so, publish the argument, and let me know.

Challenge 165, page 342: Probably none. The answer depends on whether the existence of strands can be deduced from dreams. If strands can be deduced from dreams, all of physics follows. The conjecture is that this is possible. If you find an argument against or in favour of this conjecture, let me know.

BIBLIOGRAPHY

> The only end of writing is to enable the readers better to enjoy life, or better to endure it.
>
> Samuel Johnson*

1 See the first volume of the Motion Mountain series, *Fall, Flow and Heat*, available as free download at www.motionmountain.net. Cited on pages 17 and 330.

2 See the second volume of the Motion Mountain series, *Relativity*, available as free download at www.motionmountain.net. Cited on pages 17, 18, 330, and 353.

3 See the third volume of the Motion Mountain series, *Light, Charges and Brains*, available as free download at www.motionmountain.net, as well as the mentioned fourth and fifth volumes. Cited on pages 17 and 330.

4 See the fourth and fifth volume of the Motion Mountain series, *Quantum Theory: The Smallest Change* and *Pleasure, Technology and the Stars*, available as free download at www.motionmountain.net. Cited on pages 18, 330, 331, and 353.

5 See for example, the book by Robert Laughlin, *A Different Universe: Reinventing Physics from the Botton Down* Basic Books, 2005. Of the numerous books that discuss the idea of a final theory, this is the only one worth reading, and the only one cited in this bibliography. The opinions of Laughlin are worth pondering. Cited on page 20.

6 Steven Weinberg regularly – and incorrectly – claims in interviews that the measurement problem is not solved yet. Cited on page 20.

7 Undocumented sentences to this effect are regularly attributed to Albert Einstein. Since Einstein was a pantheist, his statements on the matter are not really to be taken seriously. They were probably made – if at all – in a humorous tone. Cited on page 20.

8 For an example for the inappropriate fear of unification, see the theatre play *Die Physiker* by the Swiss author Friedrich Dürrenmatt. Several other plays and novels took over this type of disinformation. Cited on page 21.

9 Exploring the spirit of play is the subject of research of the famous National Institute for Play, founded by Stuart Brown, and found at www.nifplay.org. Cited on page 21.

10 See e.g. the 1922 lectures by Lorentz at Caltech, published as H. A. Lorentz, *Problems of Modern Physics*, edited by H. Bateman, Ginn and Company, 1927, page 99. Cited on page 25.

11 Bohr explained the indivisibilty of the quantum of action in his famous Como lecture, printed in N. Bohr, *Atomtheorie und Naturbeschreibung*, Springer, Berlin, 1931. More statements about the indivisibility of the quantum of action can be found in N. Bohr, *Atomic*

---

* This is a statement from the brilliant essay by Samuel Johnson, *Review of Soame Jenyns' "A Free Enquiry Into the Nature and Origin of Evil"*, 1757. See www.samueljohnson.com.

Physics and Human Knowledge, Science Editions, New York, 1961. For summaries of Bohr's ideas by others see Max Jammer, *The Philosophy of Quantum Mechanics*, Wiley, first edition, 1974, pp. 90–91, and John Honner, *The Description of Nature – Niels Bohr and the Philosophy of Quantum Physics*, Clarendon Press, 1987, p. 104. Cited on page 26.

12 For an overview of the quantum of action as a basis of quantum theory, see the first chapter of the fourth volume of the Motion Mountain series, Ref. 4. Cited on page 26.

13 The details of the connections with the quantum of action can be found in the volume on quantum theory. Cited on page 27.

Vol. IV, page 14

14 Minimal entropy is discussed by L. Szilard, Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen, *Zeitschrift für Physik* **53**, pp. 840–856, 1929. This classic paper can also be found in English translation in his collected works. Cited on page 27.

15 See for example A. E. Shalyt-Margolin & A. Ya. Tregubovich, Generalized uncertainty relation in thermodynamics, arxiv.org/abs/gr-qc/0307018, or J. Uffink & J. van Lith-van Dis, Thermodynamic uncertainty relations, *Foundations of Physics* **29**, pp. 655–692, 1999. Cited on page 28.

16 The first published statements of the principle of maximum force were in the second volume, *Relativity*, of the present textbook and independently in the paper by G. W. Gibbons, The maximum tension principle in general relativity, *Foundations of Physics* **32**, pp. 1891–1901, 2002, preprint at arxiv.org/abs/hep-th/0210109. See also C. Schiller, General relativity and cosmology derived from principle of maximum power or force, *International Journal of Theoretical Physics* **44**, pp. 1629–1647, 2005, preprint at arxiv.org/abs/physics/0607090. A detailed discussion is given in the volume that explores general relativity, Ref. 2. Cited on pages 29, 39, and 260.

Vol. II, page 95

17 Maximal luminosity is often mentioned in connection with gravitational wave detection; nevertheless, the general power maximum has never been mentioned before. See for example L. Ju, D. G. Blair & C. Zhao, Detection of gravitational waves, *Reports on Progress in Physics* **63**, pp. 1317–1427, 2000. See also C. Misner, K. Thorne & J. A. Wheeler, *Gravitation*, Freeman, 1973, page 980. Cited on page 29.

18 See for example Wolfgang Rindler, *Relativity – Special, General and Cosmological*, Oxford University Press, 2001, p. 70 ss, or Ray d'Inverno, *Introducing Einstein's Relativity*, Clarendon Press, 1992, p. 36 ss. Cited on page 30.

19 T. Jacobson, Thermodynamics of spacetime: the Einstein equation of state, *Physical Review Letters* **75**, pp. 1260–1263, 1995, preprint at arxiv.org/abs/gr-qc/9504004; this deep article remains fascinating to this day. The general concepts are explained, almost without formulae, in L. Smolin, On the nature of quantum fluctuations and their relation to gravitation and the principle of inertia, *Classical and Quantum Gravity* **3**, pp. 347–359, 1986. Cited on pages 30 and 259.

20 Indeterminacy relations in general relativity are discussed in C. A. Mead, Possible connection between gravitation and fundamental length, *Physical Review B* **135**, pp. 849–862, 1964. The generalized indeterminacy relation is implicit on page 852, but the issue is explained rather unclearly. Probably the author considered the result too simple to be mentioned explicitly. (That paper took 5 years to get published; comments on the story, written 37 years later, are found at C. A. Mead, Walking the Planck length through history, *Physics Today* **54**, p. 15 and p. 81, 2001, with a reply by Frank Wilczek.) See also P. K. Townsend, Small-scale structure of space-time as the origin of the gravitational constant, *Physical Review D* **15**, pp. 2795–2801, 1977, or the paper by M. -T. Jaekel & S. Renaud, Gravitational quantum limit for length measurement, *Physics Letters A* **185**, pp. 143–148, 1994. Cited on pages 32, 61, 62, 63, 65, and 122.

21 M. Kramer & al., Tests of general relativity from timing the double pulsar, preprint at arxiv.org/abs/astro-ph/060941. Cited on page 32.

22 Minimal length and minimal time intervals are discussed, for example, by G. Amelino-Camelia, Limits on the measurability of space-time distances in (the semiclassical approximation of) quantum gravity, *Modern Physics Letters A* **9**, pp. 3415–3422, 1994, preprint at arxiv.org/abs/gr-qc/9603014, and by Y. J. Ng & H. van Dam, Limit to space-time measurement, *Modern Physics Letters A* **9**, pp. 335–340, 1994. Many other authors have explored the topic. Cited on pages 34 and 62.

23 Maximal curvature, as well as area and volume quantization, are discussed in A. Ashtekar, Quantum geometry and gravity: recent advances, preprint at arxiv.org/abs/gr-qc/0112038 and in A. Ashtekar, Quantum geometry in action: big bang and black holes, preprint at arxiv.org/abs/math-ph/0202008. Cited on pages 34, 70, and 362.

24 Maximons, elementary particles of Planck mass, are discussed by A. D. Sakharov, Vacuum quantum fluctuations in curved space and the theory of gravitation, *Soviet Physics – Doklady* **12**, pp. 1040–1041, 1968. Cited on pages 36, 72, and 125.

25 Wolfgang Rindler, *Relativity – Special, General and Cosmological*, Oxford University Press, 2001, p. 230. Cited on page 38.

26 This relation was pointed out by Achim Kempf. The story is told in A. D. Sakharov, *General Relativity and Gravitation* **32**, pp. 365–367, 2000, a reprint of his paper *Doklady Akademii Nauk SSSR* **177**, pp. 70–71, 1967. Cited on page 40.

27 Several incorrect counterclaims to the entropy limit were made in R. Bousso, The holographic principle, *Review of Modern Physics* **74**, pp. 825–874, 2002, preprint at arxiv.org/abs/hep-th/0203101. However, this otherwise good review has some errors in its arguments, as explained on page 102. Bousso has changed his position in the meantime; he now accepts the entropy limit. Cited on pages 40, 44, 360, and 362.

28 Gamma ray bursts are discussed by G. Preparata, R. Ruffini & S.-S. Xue, The dyadosphere of black holes and gamma-ray bursts, *Astronomy and Astrophysics* **338**, pp. L87–L90, 1998, and C. L. Bianco, R. Ruffini & S.-S. Xue, The elementary spike produced by a pure e^+e^- pair-electromagnetic pulse from a black hole: the PEM pulse, *Astronomy and Astrophysics* **368**, pp. 377–390, 2001. Cited on page 41.

29 See for example the review in C.W.J. Beenakker & al., Quantum transport in semiconductor nanostructures, pp. 1–228, in H. Ehrenreich & D. Turnbull editors, *Solid State Physics*, volume 44, Academic Press, 1991. The prediction of a future Nobel Prize was made and conveyed to the two authors by Christoph Schiller in 1992. Cited on page 41.

30 A discussion of a different electrical indeterminacy relation, between current and charge, can be found in Y-Q. Li & B. Chen, Quantum theory for mesoscopic electronic circuits and its applications, arxiv.org/abs/cond-mat/9907171. Cited on page 41.

31 Hans C. Ohanian & Remo Ruffini, *Gravitation and Spacetime*, W.W. Norton & Co., 1994. Cited on pages 42 and 356.

32 The entropy limit for black holes is discussed by J. D. Bekenstein, Entropy bounds and black hole remnants, *Physical Review D* **49**, pp. 1912–1921, 1994. See also J. D. Bekenstein, Universal upper bound on the entropy-to-energy ratio for bounded systems, *Physical Review D* **23**, pp. 287–298, 1981. Cited on pages 44 and 135.

33 Private communications from Jos Uffink and Bernard Lavenda. No citations.

34 P. Kovtun, D. T. Son & A. O. Starinets, A viscosity bound conjecture, preprint at arxiv.org/abs/hep-th/0405231. Cited on page 45.

35 Brian Greene, *The Elegant Universe – Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*, Vintage, 2000. Cited on page 49.

36 S. Weinberg, The cosmological constant problem, *Reviews of Modern Physics* **61**, pp. 1–23, 1989. Cited on page 53.

37 Steven Weinberg, *The Quantum Theory of Fields*, Cambridge University Press, volumes I, 1995, and II, 1996. Cited on page 53.

38 The difficulties are summarised by B. S. DeWitt, Quantum field theory in curved spacetime, *Physics Reports* **19**, pp. 295–357, 1975. Cited on page 53.

39 C. W. Misner, K. S. Thorne & J. A. Wheeler, *Gravitation*, Freeman, 1973. Cited on pages 54, 56, and 62.

40 J. A. Wheeler, in *Relativity, Groups and Topology*, edited by C. DeWitt & B. S. DeWitt, Gordon and Breach, 1994. See also J. A. Wheeler, Physics at the Planck length, *International Journal of Modern Physics A* **8**, pp. 4013–4018, 1993. Cited on page 54.

41 J. L. Friedman & R. D. Sorkin, Spin 1/2 from gravity, *Physical Review Letters* **44**, pp. 1100–1103, 1980. Cited on page 54.

42 A. P. Balachandran, G. Bimonte, G. Marmo & A. Simoni, Topology change and quantum physics, *Nuclear Physics B* **446**, pp. 299–314, 1995, arxiv.org/abs/hep-th/9503046. Cited on page 54.

43 J. Ehlers, Introduction – Survey of Problems, pp. 1–10, in J. Ehlers, editor, *Sistemi gravitazionali isolati in relatività generale*, Rendiconti della scuola internazionale di fisica “Enrico Fermi”, LXVII^{o} corso, Società Italiana di Fisica/North Holland, 1979. Cited on page 54.

44 See C. Schiller, Le vide diffère-t-il de la matière? in E. Gunzig & S. Diner editors, *Le Vide – Univers du tout et du rien – Des physiciens et des philosophes s’interrogent*, Les Éditions de l’Université de Bruxelles, 1998. An older, English-language version is available as C. Schiller, Does matter differ from vacuum? arxiv.org/abs/gr-qc/9610066. Cited on pages 54, 116, 117, 122, 123, 124, 125, 126, 136, and 137.

45 See for example Richard P. Feynman, Robert B. Leighton & Matthew Sands, *The Feynman Lectures on Physics*, Addison Wesley, 1977. Cited on page 54.

46 Steven Weinberg, *Gravitation and Cosmology*, Wiley, 1972. Cited on pages 56, 60, and 62.

47 The observations of black holes at the centre of galaxies and elsewhere are summarised by R. Blandford & N. Gehrels, Revisiting the black hole, *Physics Today* **52**, June 1999. Their existence is now well established. Cited on page 56.

48 M. Planck, Über irreversible Strahlungsvorgänge, *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften zu Berlin* pp. 440–480, 1899. Today it is commonplace to use Dirac’s $\hbar = h/2\pi$ instead of Planck’s h, which Planck originally called b. Cited on pages 57 and 122.

49 The argument is given e.g. in E. P. Wigner, Relativistic invariance and quantum phenomena, *Reviews of Modern Physics* **29**, pp. 255–258, 1957. Cited on page 59.

50 The starting point for the following arguments is taken from M. Schön, Operative time definition and principal indeterminacy, arxiv.org/abs/gr-qc/9304024, and from T. Padmanabhan, Limitations on the operational definition of space-time events and quantum gravity, *Classical and Quantum Gravity* **4**, pp. L107–L113, 1987; see also Padmanabhan's earlier papers referenced there. Cited on page 59.

51 W. Heisenberg, Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, *Zeitschrift für Physik* **43**, pp. 172–198, 1927. Cited on page 59.

52 E. H. Kennard, Zur Quantenmechanik einfacher Bewegungstypen, *Zeitschrift für Physik* **44**, pp. 326–352, 1927. Cited on page 59.

53 M. G. Raymer, Uncertainty principle for joint measurement of noncommuting variables, *American Journal of Physics* **62**, pp. 986–993, 1994. Cited on page 59.

54 H. Salecker & E. P. Wigner, Quantum limitations of the measurement of space-time distances, *Physical Review* **109**, pp. 571–577, 1958. Cited on pages 60, 88, and 108.

55 E. J. Zimmerman, The macroscopic nature of space-time, *American Journal of Physics* **30**, pp. 97–105, 1962. Cited on pages 60, 88, and 108.

56 J. D. Bekenstein, Black holes and entropy, *Physical Review D* **7**, pp. 2333–2346, 1973. Cited on pages 60, 134, and 257.

57 S. W. Hawking, Particle creation by black holes, *Communications in Mathematical Physics* **43**, pp. 199–220, 1975; see also S. W. Hawking, Black hole thermodynamics, *Physical Review D* **13**, pp. 191–197, 1976. Cited on pages 60, 134, and 257.

58 On the limit for angular momentum of black holes, see Ref. 31. Cited on page 258.

59 P. Gibbs, The small scale structure of space-time: a bibliographical review, arxiv.org/abs/hep-th/9506171. Cited on pages 60 and 77.

60 M. -T. Jaekel & S. Renaud, Gravitational quantum limit for length measurement, *Physics Letters A* **185**, pp. 143–148, 1994. Cited on page 62.

61 D. V. Ahluwalia, Quantum measurement, gravitation and locality, *Physics Letters B* **339**, pp. 301–303, 1994, preprint at arxiv.org/abs/gr-qc/9308007. Cited on page 62.

62 L. Garay, Quantum gravity and minimum length, *International Journal of Modern Physics A* **10**, pp. 145–165, 1995, preprint at arxiv.org/abs/gr-qc/9403008. This paper also includes an extensive bibliography. See also R. J. Adler & D. I. Santiago, On gravity and the uncertainty principle, *Modern Physics Letters A* **14**, pp. 1371–1381, 1999, preprint at arxiv.org/abs/gr-qc/9904026. Cited on page 62.

63 C. Rovelli & L. Smolin, Discreteness of area and volume in quantum gravity, *Nuclear Physics B* **442**, pp. 593–619, 1995. R. Loll, The volume operator in discretized quantum gravity, preprint at arxiv.org/abs/gr-qc/9506014. See also C. Rovelli, Notes for a brief history of quantum gravity, preprint at arxiv.org/abs/gr-qc/0006061. Cited on page 63.

64 D. Amati, M. Ciafaloni & G. Veneziano, Superstring collisions at Planckian energies, *Physics Letters B* **197**, pp. 81–88, 1987. D. J. Gross & P. F. Mende, The high energy behavior of string scattering amplitudes, *Physics Letters B* **197**, pp. 129–134, 1987. K. Konishi, G. Paffuti & P. Provero, Minimum physical length and the generalized uncertainty principle, *Physics Letters B* **234**, pp. 276–284, 1990. P. Aspinwall, Minimum distances in non-trivial string target spaces, *Nuclear Physics B* **431**, pp. 78–96, 1994, preprint at arxiv.org/abs/hep-th/9404060. Cited on page 63.

65 M. Maggiore, A generalised uncertainty principle in quantum mechanics, *Physics Letters B* **304**, pp. 65–69, 1993. Cited on page 63.

66 A simple approach is S. Doplicher, K. Fredenhagen & J. E. Roberts, Space-time quantization induced by classical gravity, *Physics Letters B* **331**, pp. 39–44, 1994. Cited on pages 63 and 77.

67 A. Kempf, Uncertainty relation in quantum mechanics with quantum group symmetry, *Journal of Mathematical Physics* 35, pp. 4483–4496, 1994. A. Kempf, Quantum groups and quantum field theory with nonzero minimal uncertainties in positions and momenta, *Czechoslovak Journal of Physics* 44, pp. 1041–1048, 1994. Cited on page 63.

68 E. J. Hellund & K. Tanaka, Quantized space-time, *Physical Review* **94**, pp. 192–195, 1954. Cited on page 63.

69 The impossibility of determining temporal ordering in quantum theory is discussed by J. Oppenheimer, B. Reznik & W. G. Unruh, Temporal ordering in quantum mechanics, *Journal of Physics A* 35, pp. 7641–7652, 2001, preprint at arxiv.org/abs/quant-ph/0003130. Cited on page 61.

70 A. Peres & N. Rosen, Quantum limitations on the measurement of gravitational fields, *Physical Review* 118, pp. 335–336, 1960. Cited on page 65.

71 It is the first definition in Euclid's *Elements*, *c.* 300 bce. For an English translation see T. Heath, *The Thirteen Books of the Elements*, Dover, 1969. Cited on page 66.

72 A beautiful description of the Banach–Tarski paradox is the one by Ian Stewart, Paradox of the spheres, *New Scientist*, 14 January 1995, pp. 28–31. Cited on page 67.

73 This intriguing extract from a letter by Einstein was made widely known by John J. Stachel, in his paper The Other Einstein: Einstein Contra Field Theory, that is best found in his book *Einstein from 'B' to 'Z'*, Birkhäuser, 2002. The German original of the letter is found in Robert Schulmann, A. J. Knox, Michel Janssen & József Illy, *The Collected Papers of Albert Einstein, Volume 8A – The Berlin Years: Correspondence, 1914–1917*, letter 299, Princeton University Press, 1998. Barbara Wolff helped in clarifying several details in the German original. Cited on page 64.

74 H. S. Snyder, Quantized space-time, *Physical Review* 71, pp. 38–41, 1947. H. S. Snyder, The electromagnetic field in quantized space-time, *Physical Review* 72, pp. 68–74, 1947. A. Schild, Discrete space-time and integral Lorentz transformations, *Physical Review* 73, pp. 414–415, 1948. E. L. Hill, Relativistic theory of discrete momentum space and discrete space-time, *Physical Review* **100**, pp. 1780–1783, 1950. H. T. Flint, The quantization of space-time, *Physical Review* 74, pp. 209–210, 1948. A. Das, Cellular space-time and quantum field theory, *Il Nuovo Cimento* 18, pp. 482–504, 1960. Cited on page 68.

75 D. Finkelstein, 'Superconducting' causal nets, *International Journal of Theoretical Physics* 27, pp. 473–519, 1985. Cited on page 68.

76 N. H. Christ, R. Friedberg & T. D. Lee, Random lattice field theory: general formulation, *Nuclear Physics B* **202**, pp. 89–125, 1982. G. 't Hooft, Quantum field theory for elementary particles – is quantum field theory a theory?, *Physics Reports* **104**, pp. 129–142, 1984. Cited on page 69.

77 For a discussion, see R. Sorabji, *Time, Creation and the Continuum: Theories in Antiquity and the Early Middle Ages*, Duckworth, 1983. Cited on page 69.

78 See, for example, L. Bombelli, J. Lee, D. Meyer & R. D. Sorkin, Space-time as a causal set, *Physical Review Letters* **59**, pp. 521–524, 1987. G. Brightwell & R. Gregory, Structure of random space-time, *Physical Review Letters* **66**, pp. 260–263, 1991. Cited on page 69.

79 The false belief that particles like quarks or electrons are composite is slow to die out. See for example: S. Fredriksson, Preon prophecies by the standard model, arxiv.org/abs/hep-ph/0309213. Preon models gained popularity in the 1970s and 1980s, in particular through the papers by J. C. Pati & A. Salam, Lepton number as the fourth "color", *Physical Review D* **10**, pp. 275–289, 1974, H. Harari, A schematic model of quarks and leptons, *Physics Letters B* **86**, pp. 83–86, 1979, M. A. Shupe, A composite model of leptons and quarks, *Physics Letters B* **86**, pp. 87–92, 1979, and H. Fritzsch & G. Mandelbaum, Weak interactions as manifestations of the substructure of leptons and quarks, *Physics Letters B* **102**, pp. 319–322, 1981. Cited on page 70.

80 C. Wolf, Upper limit for the mass of an elementary particle due to discrete time quantum mechanics, *Il Nuovo Cimento B* **109**, pp. 213–218, 1994. Cited on page 73.

81 N. F. Ramsey & A. Weis, Suche nach permanenten elektrischen Dipolmomenten: ein Test der Zeitumkehrinvarianz, *Physikalische Blätter* **52**, pp. 859–863, 1996. The paper by E. D. Commins, S. B. Ross, D. DeMille & B. C. Regan, Improved experimental limit on the electric dipole moment of the electron, *Physical Review A* **50**, pp. 2960–2977, 1994, gives an upper experimental limit to the dipole moment of the electron of $3.4 \cdot 10^{-29}\, e\,\mathrm{m}$. W. Bernreuther & M. Suzuki, The electric dipole moment of the electron, *Reviews of Modern Physics* **63**, pp. 313–340, 1991. See also the musings in Hans Dehmelt, Is the electron a composite particle?, *Hyperfine Interactions* **81**, pp. 1–3, 1993. Cited on page 71.

82 K. Akama, T. Hattori & K. Katsuura, Naturalness bounds on dipole moments from new physics, preprint at arxiv.org/abs/hep-ph/0111238. Cited on page 71.

83 W. G. Unruh, Notes on black hole evaporation, *Physical Review D* **14**, pp. 870–875, 1976. W. G. Unruh & R. M. Wald, What happens when an accelerating observer detects a Rindler particle, *Physical Review D* **29**, pp. 1047–1056, 1984. Cited on page 75.

84 The first example was J. Magueijo & L. Smolin, Lorentz invariance with an invariant energy scale, *Physical Review Letters* **88**, p. 190403, 2002, or arxiv.org/abs/hep-th/0112090. They propose a modification of the mass energy relation of the kind

$$E = \frac{\gamma m c^2}{1 + \frac{\gamma m c^2}{E_{\mathrm{Pl}}}} \quad \text{and} \quad p = \frac{\gamma m v}{1 + \frac{\gamma m c^2}{E_{\mathrm{Pl}}}} \,. \tag{214}$$

Another, similar approach of recent years, with a different proposal, is called ‘doubly special relativity’. A recent summary is G. Amelino-Camelia, Doubly-special relativity: first results and key open problems, *International Journal of Modern Physics* **11**, pp. 1643–1669, 2002, preprint at arxiv.org/abs/gr-qc/0210063. The paper shows how conceptual problems hinder the advance of the field. Another such discussion R. Aloisio, A. Galante, A. F. Grillo, E. Luzio & F. Méndez, Approaching space-time through velocity in doubly special relativity, preprint at arxiv.org/abs/gr-qc/0410020. The lesson from these attempts is simple: special relativity *cannot* be modified to include a limit energy without also including general relativity and quantum theory. Cited on pages 77 and 248.

85 W. Jauch, Heisenberg’s uncertainty relation and thermal vibrations in crystals, *American Journal of Physics* **61**, pp. 929–932, 1993. Cited on page 77.

86 H. D. Zeh, On the interpretation of measurement in quantum theory, *Foundations of Physics* **1**, pp. 69–76, 1970. Cited on page 78.

87 See Y. J. Ng, W. A. Christiansen & H. van Dam, Probing Planck-scale physics with extragalactic sources?, *Astrophysical Journal* **591**, pp. L87–L90, 2003, preprint at arxiv.org/abs/astro-ph/0302372; D. H. Coule, Planck scale still safe from stellar images, *Classical and Quantum Gravity* **20**, pp. 3107–3112, 2003, preprint at arxiv.org/abs/astro-ph/0302333. Negative experimental results (and not always correct calculations) are found in R. Lieu & L. Hillman, The phase coherence of light from extragalactic sources – direct evidence against first order Planck scale fluctuations in time and space, *Astrophysical Journal* **585**, pp. L77–L80, 2003, and R. Ragazzoni, M. Turatto & W. Gaessler, The lack of observational evidence for the quantum structure of spacetime at Planck scales, *Astrophysical Journal* 587, pp. L1–L4, 2003. Cited on pages 81 and 82.

88 B. E. Schaefer, Severe limits on variations of the speed of light with frequency, *Physical Review Letters* 82, pp. 4964–4966, 21 June 1999. Cited on page 81.

89 A.A. Abdo & al., Testing Einstein's special relativity with Fermi's short hard γ-ray burst GRB090510, preprint at arxiv.org/0908.1832. Cited on page 81.

90 G. Amelino-Camelia, J. Ellis, N. E. Mavromatos, D. V. Nanopoulos & S. Sakar, Potential sensitivity of gamma-ray-burster observations to wave dispersion in vacuo, *Nature* 393, pp. 763–765, 1998, preprint at arxiv.org/abs/astro-ph/9712103. Cited on page 81.

91 G. Amelino-Camelia, Phenomenological description of space-time foam, preprint at arxiv.org/abs/gr-qc/0104005. The paper includes a clearly written overview of present experimental approaches to detecting quantum gravity effects. See also his update G. Amelino-Camelia, Quantum-gravity phenomenology: status and prospects, preprint at arxiv.org/abs/gr-qc/0204051. Cited on pages 81 and 82.

92 G. Amelino-Camelia, An interferometric gravitational wave detector as a quantum gravity apparatus, *Nature* 398, pp. 216–218, 1999, preprint at arxiv.org/abs/gr-qc/9808029. Cited on page 81.

93 F. Karolyhazy, Gravitation and quantum mechanics of macroscopic objects, *Il Nuovo Cimento* A42, pp. 390–402, 1966. Y. J. Ng & H. van Dam, Limit to space-time measurement, *Modern Physics Letters A* 9, pp. 335–340, 1994. Y. J. Ng & H. van Dam, *Modern Physics Letters A* Remarks on gravitational sources, 10, pp. 2801–2808, 1995. The discussion is neatly summarised in Y. J. Ng & H. van Dam, Comment on 'Uncertainty in measurements of distance', preprint at arxiv.org/abs/gr-qc/0209021. See also Y. J. Ng, Spacetime foam, preprint at arxiv.org/abs/gr-qc/0201022. Cited on pages 81 and 88.

94 L. J. Garay, Spacetime foam as a quantum thermal bath, *Physics Review Letters* 80, pp. 2508–2511, 1998, preprint at arxiv.org/abs/gr-qc/9801024. Cited on pages 81 and 82.

95 G. Amelino-Camelia & T. Piran, Planck-scale deformation of Lorentz symmetry as a solution to the UHECR and the TeV-γ paradoxes, preprint at arxiv.org/astro-ph/0008107, 2000. Cited on page 82.

96 R. P. Woodard, How far are we from the quantum theory of gravity?, preprint at arxiv.org/abs/0907.4238. For a different point of view, see L. Smolin, Generic predictions of quantum theories of gravity, preprint at arxiv.org/abs/hep-th/0605052. Cited on pages 83 and 266.

97 See the lucid discussion by G. F. R. Ellis & T. Rothman, Lost horizons, *American Journal of Physics* 61, pp. 883–893, 1993. Cited on pages 88, 92, and 93.

98 See, for example, the Hollywood film *Contact*, based on the book by Carl Sagan, *Contact*, 1985. Cited on page 93.

99 See, for example, the international bestseller by Stephen Hawking, *A Brief History of Time – From the Big Bang to Black Holes*, 1988. Cited on page 95.

100 L. Rosenfeld, Quantentheorie und Gravitation, in H.-J. Treder, editor, *Entstehung, Entwicklung und Perspektiven der Einsteinschen Gravitationstheorie*, Springer Verlag, 1966. Cited on page 98.

101 Holography in high-energy physics is connected with the work of 't Hooft and Susskind. See for example G. 't Hooft, Dimensional reduction in quantum gravity, pp. 284–296, in A. Ali, J. Ellis & S. Randjbar-Daemi, *Salaamfeest*, 1993, or the much-cited paper by L. Susskind, The world as a hologram, *Journal of Mathematical Physics* **36**, pp. 6377–6396, 1995, or arxiv.org/abs/hep-th/9409089. A good modern overview is Ref. 27. Cited on pages 100 and 135.

102 D. Bohm & B. J. Hiley, On the intuitive understanding of nonlocality as implied by quantum theory, *Foundations of Physics* **5**, pp. 93–109, 1975. Cited on page 101.

103 S. Lloyd, Computational capacity of the universe, *Physical Review Letters* **88**, p. 237901, 2002. Cited on page 102.

104 Gottfried Wilhelm Leibniz, *La Monadologie*, 1714. Written in French, it is available freely at www.uqac.uquebec.ca/zone30/Classiques_des_sciences_sociales and in various other languages on other websites. Cited on page 102.

105 See, for example, H. Wussing & P. S. Alexandrov editors, *Die Hilbertschen Probleme*, Akademische Verlagsgesellschaft Geest & Portig, 1983, or Ben H. Yandell, *The Honours Class: Hilbert's Problems and their Solvers*, A.K. Peters, 2002. Cited on page 103.

106 A large part of the study of dualities in string and M theory can be seen as investigations into the detailed consequences of extremal identity. For a review of dualities, see P. C. Argyres, Dualities in supersymmetric field theories, *Nuclear Physics Proceedings Supplement* **61**, pp. 149–157, 1998, preprint available at arxiv.org/abs/hep-th/9705076. A classical version of duality is discussed by M. C. B. Abdalla, A. L. Gadelka & I. V. Vancea, Duality between coordinates and the Dirac field, arxiv.org/abs/hep-th/0002217. Cited on page 107.

107 If you are more interested in sexuality than in love, one of the best books to read is the fascinating text by Olivia Judson, *Dr. Tatjana's Sex Advice to all Creation*, Metropolitan Books, 2002. Cited on page 109.

108 The consequences of memory loss in this case are already told by Voltaire, *Aventure de la mémoire*, 1775. Cited on page 111.

109 A picture of objects in a red-hot oven and at room temperature is shown in C. H. Bennett, Demons, Engines and the Second Law, *Scientific American* **255**, pp. 108–117, November 1987. Another is shown in the third volume of the *Motion Mountain* series. Cited on page 111.

110 An introduction to how the sense of time arises as a result of signals in the ganglia and in the temporal lobe of the brain is found in R. B. Ivry & R. Spencer, The neural representation of time, *Current Opinion in Neurobiology* **14**, pp. 225–232, 2004. Cited on page 112.

111 This famous statement is found at the beginning of chapter XI, 'The Physical Universe', in Arthur Eddington, *The Philosophy of Physical Science*, Cambridge, 1939. Cited on pages 113 and 131.

112 See Chapter 7 on brain and language for details and references. Cited on page 114.

113 A similar point of view, often called monism, was proposed by Baruch Spinoza, *Ethics Demonstrated in Geometrical Order*, 1677, originally in Latin; an affordable French edition is B. Spinoza, *L'Ethique*, Folio-Gallimard, 1954. For a discussion of his ideas, especially his monism, see Don Garret editor, *The Cambridge Companion to Spinoza*, Cambridge University Press, 1996, or any general text on the history of philosophy. Cited on page 84.

114 See L. Susskind & J. Uglum, Black holes, interactions, and strings, arxiv.org/abs/hep-th/9410074, or L. Susskind, String theory and the principle of black hole complementarity, *Physical Review Letters* **71**, pp. 2367–2368, 1993, and M. Karliner, I. Klebanov & L. Susskind, Size and shape of strings, *International Journal of Modern Physics A* **3**, pp. 1981–1996, 1988, as well as L. Susskind, Structure of hadrons implied by duality, *Physical Review D* 1, pp. 1182–1186, 1970. Cited on pages 121 and 135.

115 P. Facchi & S. Pascazio, Quantum Zeno and inverse quantum Zeno effects, pp. 147–217, in E. Wolf editor, *Progress in Optics*, 42, 2001. Cited on page 125.

116 Aristotle, *Of Generation and Corruption*, book I, part 2. See Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 427, 1991. Cited on page 125.

117 See for example the speculative model of vacuum as composed of Planck-size spheres proposed by F. Winterberg, *Zeitschrift für Naturforschung* 52a, p. 183, 1997. Cited on page 126.

118 The Greek salt-and-water argument and the fish argument are given by Lucrece, in full Titus Lucretius Carus, *De natura rerum*, *c.* 60 bce. Cited on pages 127 and 141.

119 J. H. Schwarz, The second superstring revolution, Colloquium-level lecture presented at the Sakharov Conference in Moscow, May 1996, arxiv.org/abs/hep-th/9607067. Cited on pages 128 and 129.

120 Simplicius, *Commentary on the Physics of Aristotle*, 140, 34. This text is cited in Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 379, 1991. Cited on page 128.

121 D. Olive & C. Montonen, Magnetic monopoles as gauge particles, *Physics Letters* 72B, pp. 117–120, 1977. Cited on page 129.

122 A famous fragment from Diogenes Laertius quotes Democritus as follows: ‘By convention hot, by convention cold, but in reality, atoms and void; and also in reality we know nothing, since truth is at the bottom.’ Cited on page 130.

123 Plato, *Parmenides*, *c.* 370 bce. It has been translated into most languages. Reading it aloud, like a song, is a beautiful experience. A pale reflection of these ideas is Bohm’s concept of ‘unbroken wholeness’. Cited on page 131.

124 P. Gibbs, Event-symmetric physics, arxiv.org/abs/hep-th/9505089; see also his website www.weburbia.com/pg/contents.htm. Cited on page 132.

125 J. B. Hartle & S. W. Hawking, Path integral derivation of black hole radiance, *Physical Review D* 13, pp. 2188–2203, 1976. See also A. Strominger & C. Vafa, Microscopic origin of Bekenstein–Hawking entropy, *Physics Letters B* 379, pp. 99–104, 1996, preprint at arxiv.org/abs/hep-th/9601029. For another derivation of black hole entropy, see G. T. Horowitz & J. Polchinski, A correspondence principle for black holes and strings, *Physical Review D* 55, pp. 6189–6197, 1997, or arxiv.org/abs/hep-th/9612146. Cited on pages 134 and 143.

126 J. Maddox, When entropy does not seem extensive, *Nature* 365, p. 103, 1993. The issue is now explored in all textbooks discussing black holes. John Maddox (1925–1999) was famous for being one of the few people who was knowledgeable in most natural sciences. Cited on page 135.

127 L. Bombelli, R. K. Koul, J. Lee & R. D. Sorkin, Quantum source of entropy of black holes, *Physical Review D* 34, pp. 373–383, 1986. Cited on page 135.

128 The analogy between polymers and black holes is due to G. Weber, Thermodynamics at boundaries, *Nature* 365, p. 792, 1993. Cited on page 135.

129 Pierre-Gilles de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, 1979. Cited on page 135.

130 See for example S. Majid, Introduction to braided geometry and q-Minkowski space, preprint at arxiv.org/abs/hep-th/9410241, or S. Majid, Duality principle and braided geometry, preprint at arxiv.org/abs/hep-th/940957. Cited on pages 136 and 137.

131 The relation between spin and statistics has been studied recently by M. V. Berry & J. M. Robbins, Quantum indistinguishability: spin–statistics without relativity or field theory?, in R. C. Hilborn & G. M. Tino editors, *Spin–Statistics Connection and Commutation Relations*, American Institute of Physics, 2000. Cited on page 138.

132 On the present record, see en.wikipedia.org/wiki/Ultra-high-energy_cosmic_ray and fr.wkipedia.org/wiki/Zetta-particule. Cited on page 140.

133 P. F. Mende, String theory at short distance and the principle of equivalence, preprint at arxiv.org/abs/hep-th/9210001. Cited on page 140.

134 A. Gregori, Entropy, string theory, and our world, preprint at arxiv.org/abs/hep-th/0207195. Cited on pages 139 and 141.

135 An example is given by A. A. Slavnov, Fermi–Bose duality via extra dimension, preprint at arxiv.org/abs/hep-th/9512101. See also the standard work by Michael Stone editor, *Bosonization*, World Scientific, 1994. Cited on page 141.

136 A popular account making the point is the bestseller by the US-American string theorist Brian Greene, *The Elegant Universe – Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*, Jonathan Cape 1999. No citations.

137 A weave model of space-time appears in certain approaches to quantum gravity, such as Ref. 23. On a slightly different topic, see also S. A. Major, A spin network primer, arxiv.org/abs/gr-qc/9905020. Cited on page 141.

138 L. Smolin & Y. Wan, Propagation and interaction of chiral states in quantum gravity, preprint at arxiv.org/abs/0710.1548, and references therein. Cited on page 141.

139 A good introduction into his work is the paper D. Kreimer, New mathematical structures in renormalisable quantum field theories, *Annals of Physics* 303, pp. 179–202, 2003, erratum ibid. 305, p. 79, 2003, preprint at arxiv.org/abs/hep-th/0211136. Cited on page 141.

140 Introductions to holography include E. Alvarez, J. Conde & L. Hernandez, Rudiments of holography, arxiv.org/abs/hep-th/0205075, and Ref. 27. The importance of holography in theoretical high-energy physics was underlined by the discovery of J. Maldacena, The large N limit of superconformal field theories and supergravity, arxiv.org/abs/hep-th/9711200. Cited on page 142.

141 X. -G. Wen, From new states of matter to a unification of light and electrons, preprint at arxiv.org/abs/0508020. Cited on page 142.

142 J. S. Avrin, A visualizable representation of the elementary particles, *Journal of Knot Theory and Its Ramifications* 14, pp. 131–176, 2005. Cited on pages 142 and 301.

143 The well-known ribbon model is presented in S. Bilson-Thompson, A topological model of composite preons, preprint at arxiv.org/hep-ph/0503213; S. Bilson-Thompson, F. Markopoulou & L. Smolin, Quantum gravity and the standard model, preprint at arxiv.org/hep-th/0603022; S. Bilson-Thompson, J. Hackett, L. Kauffman & L. Smolin, Particle identifications from symmetries of braided ribbon network invariants, preprint at arxiv.org/abs/0804.0037; S. Bilson-Thompson, J. Hackett & L. Kauffman, Particle topology, braids, and braided belts, preprint at arxiv.org/abs/0903.1376. Cited on pages 142, 161, and 301.

144 R. J. Finkelstein, A field theory of knotted solitons, preprint at arxiv.org/abs/hep-th/0701124. See also R. J. Finkelstein, Trefoil solitons, elementary fermions, and $SU_q(2)$, preprint at arxiv.org/abs/hep-th/0602098, R. J. Finkelstein & A. C. Cadavid, Masses and interactions of q-fermionic knots, preprint at arxiv.org/abs/hep-th/0507022, and R. J. Finkelstein, A knot model suggested by the standard electroweak theory, preprint at arxiv.org/abs/hep-th/0408218. Cited on pages 142 and 301.

145 Louis H. Kauffman, *Knots and Physics*, World Scientific, 1991. This is a wonderful book. Cited on pages 142 and 245.

146 S. K. Ng, On a knot model of the π^+ meson, preprint at arxiv.org/abs/hep-th/0210024, and S. K. Ng, On a classification of mesons, preprint at arxiv.org/abs/hep-ph/0212334. Cited on pages 142 and 301.

147 String cosmology is a pastime for many. Examples include N. E. Mavromatos, String cosmology, arxiv.org/abs/hep-th/0111275, and N. G. Sanchez, New developments in string gravity and string cosmology – a summary report, arxiv.org/abs/hep-th/0209016. Cited on page 139.

148 Searches for background-free approaches are described by E. Witten, Quantum background independence in string theory, arxiv.org/abs/hep-th/9306122 and E. Witten, On background-independent open string theory, arxiv.org/abs/hep-th/9208027. Cited on page 146.

149 For a good introduction to string theory, see the lectures by B. Zwiebach, String theory for pedestrians, agenda.cern.ch/fullAgenda.php?ida=a063319. Cited on pages 142 and 302.

150 See A. Sen, An introduction to duality symmetries in string theory, in *Les Houches Summer School: Unity from Duality: Gravity, Gauge Theory and Strings (Les Houches, France, 2001)*, Springer Verlag, **76**, pp. 241–322, 2002. Cited on page 142.

151 Brian Greene regularly uses the name *string conjecture*. For example, he did so in a podium discussion at TED in 2009; the video of the podium discussion can be downloaded at www.ted.org. Cited on page 143.

152 L. Susskind, Some speculations about black hole entropy in string theory, arxiv.org/abs/hep-th/9309145. G. T. Horowitz & J. Polchinski, A correspondence principle for black holes and strings, *Physical Review D* 55, pp. 6189–6197, 1997, preprint at arxiv.org/abs/hep-th/9612146. Cited on page 143.

153 F. Wilczek, Getting its from bits, *Nature* 397, pp. 303–306, 1999. Cited on page 144.

154 M. R. Douglas, Understanding the landscape, preprint at arxiv.org/abs/hep-th/0602266; his earlier papers also make the point. Cited on page 144.

155 The difficulties of string theory are discussed in the well-known internet blog by Peter Woit, *Not even wrong*, at www.math.columbia.edu/~woit/blog. Several Nobel Prize winners for particle physics dismiss string theory: Martin Veltman, Sheldon Glashow, Burton Richter, Richard Feynman and since 2009 also Steven Weinberg are among those who did so publicly. Cited on pages 145 and 161.

156 S. Carlip, The small scale structure of spacetime, preprint at arxiv.org/abs/1009.1136. This paper deduces the existence of fluctuating lines in vacuum from various different arguments that are completely independent of the strand model. Cited on pages 157 and 264.

157 David Deutsch says that any good explanation must be ‘hard to vary’. This must also apply a unified model, as it claims to explain everything that is observed. See D. Deutsch, A new way to explain explanation, video talk at www.ted.org. Cited on pages 160 and 334.

158 L. Bombelli, J. Lee, D. Meyer & R. Sorkin, Space-time as a causal set, *Physical Review Letters* 59, pp. 521–524, 1987. See also the review by J. Henson, The causal set approach to quantum gravity, preprint at arxiv.org/abs/gr-qc/0601121. Cited on pages 161 and 264.

159 D. Finkelstein, Homotopy approach to quantum gravity, *International Journal of Theoretical Physics* 47, pp. 534–552, 2008. Cited on pages 161 and 264.

160 L. H. Kauffman & S. J. Lomonaco, Quantum knots, preprint at arxiv.org/abs/quant-ph/0403228. See also S. J. Lomonaco & L. H. Kauffman, Quantum knots and mosaics, preprint at arxiv.org/abs/quant-ph/0805.0339. Cited on page 161.

161 Immanuel Kant, *Critik der reinen Vernunft*, 1781, is a famous and long book that every philospher pretends to have read. Cited on page 163.

162 The literature on circularity is rare. For two interesting exceptions, see L. H. Kauffman, *Knot logic* downloadable from www2.math.uic.edu/~kauffman, and L. H. Kauffman, Reflexivity and eigenform, *Constructivist Foundations* 4, pp. 121–137, 2009. Cited on page 164.

163 Information on the belt trick is scattered across many books and few papers. The best source of information on this topic are websites. For belt trick visualizations see www.evl.uic.edu/hypercomplex/html/dirac.html, www.evl.uic.edu/hypercomplex/html/handshake.html, or www.gregegan.net/APPLETS/21/21.html. For an excellent literature summary and more movies, see www.math.utah.edu/~palais/links.html. None of these sites or the cited references seem to mention that there are *two* ways to perform the belt trick; this seems to be hidden knowledge. In September 2009, at my proposal, Greg Egan changed his applet to show both versions of the belt trick. Cited on pages 169 and 170.

164 D. Bohm, R. Schiller & J. Tiomno, A causal interpretation of the Pauli equation (A), *Supplementi al Nuovo Cimento* **1**, pp. 48 – 66, 1955, and D. Bohm & R. Schiller, A causal interpretation of the Pauli equation (B), *Supplementi al Nuovo Cimento* **1**, pp. 67–91, 1955. The authors explore an unusual way to interpret the wavefunction, which is of little interest here; but doing so, they give and explore the description of Pauli spinors in terms of Euler angles. Cited on page 186.

165 This algebraic transformation is shown in all textbooks that treat the Pauli equation. It can also be checked by writing the two equations out component by component. Cited on page 186.

166 R. P. Feynman, *QED – The Strange Theory of Light and Matter*, Princeton University Press 1988. Cited on pages 187 and 201.

167 This point was made in a lucid seminar at the University of Amsterdam in the 1990s. A single-valued many-particle wave function in $\mathbb{R}^{3N}$ dimensions is mathematically equivalent to a N-valued wave function in three-dimensional space $\mathbb{R}^3$. Cited on page 189.

168 A. Aspect, J. Dalibard & G. Roger, Experimental tests of Bell's inequalities using time-varying analyzers, *Physical Review Letters* **49**, pp. 1804–1807, 1982, Cited on page 191.

169 L. Kauffman, New invariants of knot theory, *American Mathematical Monthly* **95**, pp. 195–242, 1987. See also the image at the start of chapter 6 of Louis H. Kauffman, *On Knots*, Princeton University Press, 1987. Cited on page 193.

170 S. Kochen & E. P. Specker, The problem of hidden variables in quantum mechanics, 17, pp. 59–87, 1967. Cited on page 193.

171 This issue is explained in any text on quantum electrodynamics. Cited on page 197.

172 J.-M. Lévy-Leblond, Nonrelativitic particles and wave equations, *Communications in Mathematical Physics* **6**, pp. 286–311, 1967. See also A. Galindo & C. Sánchez del Río, Intrinsic magnetic moment as a nonrelativistic phenomenon, *American Journal of Physics* **29**, pp. 582–584, 1961, and V. I. Fushchich, A. G. Nikitin & V. A. Salogub, On the non-relativistic motion equations in the Hamiltonian form, *Reports on Mathematical Physics* **13**, pp. 175–185, 1978. Cited on page 198.

173 L. Lerner, Derivation of the Dirac equation from a relativistic representation of spin, *European Journal of Phyics* **17**, pp. 172–175, 1996. Cited on pages 198 and 200.

174 E. P. Battey-Pratt & T. J. Racey, Geometric model for fundamental particles, *International Journal of Theoretical Physics* **19**, pp. 437–475, 1980. Cited on pages 199, 200, and 373.

175 A. Abraham, Prinzipien der Dynamik des Elektrons, *Annalen der Physik* **10**, pp. 105–179, 1903, J. Frenkel, Die Elektrodynamik des rotierenden Elektrons, *Zeitschrift für Physik* **37**, pp. 243–262, 1926, L. H. Thomas, The motion of a spinning electron, *Nature* April 10, p. 514, 1926, and L. H. Thomas, The kinematics of an electron with an axis, *Philosophical Magazine* **3**, pp. 1–22, 1927. See also W. E. Baylis, Surprising symmetries in relativistic charge dynamics, preprint at arxiv.org/abs/physics/0410197. See also W. E. Baylis, Quantum/classical interface: a geometric approach from the classical side, pp. 127–154 and W. E. Baylis, Geometry of paravector space with applications to relativistic physics, pp. 363–387 in *Computational Noncommutative Algebra and Applications*, Proceedings of the NATO Advanced Study Institute, NATO Science Series II, vol. 136, ed. J. Byrnes, Kluwer Academic 2004. W. E. Baylis, R. Cabrera & D. Keselica, Quantum/classical interface: fermion spin, preprint at arxiv.org/abs/0710.3144. D. Hestenes, Zitterbewegung Modelling, *Foundations of Physics* **23**, pp. 365–386, 1993. D. Hestenes, Zitterbewegung in quantum mechanics – a research program, preprint at arxiv.org/abs/0802.2728. See also D. Hestenes, Reading the electron clock, preprint at arxiv.org/abs/0802.3227 and his webpage modelingnts.la.asu.edu/html/GAinQM.html. A. Loinger & A. Sparzani, Dirac equation without Dirac matrices, *Il Nuovo Cimento* **39**, pp. 1140–1145, 1965. D. Bohm, P. Hillion, T. Takabayasi & J. -P. Vigier, Relativistic rotators and bilocal theory, *Progress of Theoretical Physics* **23**, pp. 496–511, 1960. A. Challinor, A. Lasenby, S. Gill & C. Doran, A relativistic, causal account of a spin measurement, *Physics Letters A* **218**, pp. 128–138, 1996. E. Santamato, The role of Dirac equation in the classical mechanics of the relativistic top, preprint at arxiv.org/abs/0808.3237. Cited on page 201.

176 The concept of Zitterbewegung was formulated in E. Schrödinger, Über die kräftefreie Bewegung in der relativistischen Quantenmechanik, *Berliner Berichte* pp. 418–428, 1930, and Zur Quantendynamik des Elektrons, *Berliner Berichte* pp. 63–72, 1931. Numerous subsequent papers discuss these publications. Cited on page 201.

177 See for example the book by Martin Rivas, *Kinematic Theory of Spinning Particles*, Springer, 2001. Cited on page 202.

178 The basic papers in the field of stochastic quantization are W. Weizel, Ableitung der Quantentheorie aus einem klassischen, kausal determinierten Modell, *Zeitschrift für Physik A* **134**, pp. 264–285, 1953, W. Weizel, Ableitung der Quantentheorie aus einem klassischen Modell – II, *Zeitschrift für Physik A* **135**, pp. 270–273, 1954, W. Weizel, Ableitung der quantenmechanischen Wellengleichung des Mehrteilchensystems aus einem klassischen Modell, *Zeitschrift für Physik A* **136**, pp. 582–604, 1954. This work was taken up by E. Nelson, Derivation of the Schrödinger equation from Newtonian mechanics, *Physical Review* **150**, pp. 1079–1085, 1969, and in Edward Nelson, *Quantum Fluctuations*, Princeton University Press 1985, also downloadable at www.math.princeton.edu/~nelson/books.html, and the book Edward Nelson, *Stochastic Quantization*, Princeton University Press 1985, See also L. Fritsche & M. Haugk, A new look at the derivation of the Schrödinger equation from Newtonian mechanics, *Annalen der Physik* **12**, pp. 371–402, 2003. A summary of Nelson's approach is also given in F. Markopoulou & L. Smolin, Quantum theory from quantum gravity, *Physical Review D* **70**, p. 124029, 2004, preprint at www.arxiv.org/abs/gr-qc/0311059. See also the important criticism by T. C. Wallstrom, Inequivalence between the Schrödinger equation and the Madelung hydrodynamic equation, *Physical Review A* **49**, pp. 1613–1617, 1994, and T. C. Wallstrom, The stochastic mechanics of the Pauli equation, *Transactions of the American Mathematical Society* **318**, pp. 749–762, 1990. A proposed answer is L. Smolin, Could quantum mechanics be an approximation to another theory?, preprint at arxiv.org/quant-ph/abs/0609109. See also S. K. Srinivasan & E. C. G. Sudarshan, A direct derivation of the Dirac equation via quaternion measures, *Journal of Physics A* **29**, pp. 5181–5186, 1996. Cited on page 202.

179 For such an attempt, see the proposal by M. Rainer, Resolution of simple singularities yielding particle symmetries in space-time, *Journal of Mathematical Physics* **35**, pp. 646–655, 1994. Cited on page 207.

180 C. Schiller, Deducing the three gauge ineractions from featureless strands, preprint at arxiv.org/abs/0905.3905. Cited on pages 207 and 209.

181 G. T. Horowitz & J. Polchinski, Gauge/gravity duality, preprint at arxiv.org/abs/gr-qc/0602037. Note also the statement in the introduction that a graviton might be a composite of two spin-1 bosons, which is somewhat reproduced by the strand model of the graviton. A more concrete approach to gauge–gravity duality is made by M. van Raamsdonk, Building up spacetime with quantum entanglement, preprint at arxiv.org/1005.3035. This approach to gauge–gravity duality is close to that of the strand model. Cited on page 208.

182 K. Reidemeister, Elementare Begründung der Knotentheorie, *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg* **5**, pp. 24–32, 1926.. Cited on pages 209 and 245.

183 For some of the background on this topic, see F. Wilczek & A. Zee, Appearance of gauge structures in simple dynamical systems, *Physical Review Letters* **52**, pp. 2111–2114, 1984, A. Shapere & F. Wilczek, Self-propulsion at low Reynolds number, *Physical Review Letters* **58**, pp. 2051–2054, 1987, and A. Shapere & F. Wilczek, Gauge kinematics of deformable bodies, *American Journal of Physics* **57**, pp. 514–518, 1989. Cited on page 244.

184 R. Britto, F. Cachazo, B. Feng & E. Witten, Direct proof of tree-level recursion relation in Yang–Mills theory, preprint at arxiv.org/abs/hep-th/0501052. Cited on page 244.

185 D. V. Ahluwalia-Khalilova, Operational indistinguishability of double special relativity from special relativity, *Classical and Quantum Gravity* **22**, pp. 1433–1450, 2005, preprint at arxiv.org/abs/gr-qc/0212128; see also N. Jafari & A. Shariati, Doubly special relativity: a new relativity or not?, preprint at arxiv.org/abs/gr-qc/0602075. Cited on page 248.

186 E. Verlinde, On the origin of gravity and the laws of Newton, preprint at arxiv.org/abs/1001.0785. Cited on page 250.

187 G.-L. Lesage, Lucrèce Newtonien, *Nouveaux mémoires de l'Académie Royale des Sciences et Belles Lettres* pp. 404–431, 1747, or www3.bbaw.de/bibliothek/digital/struktur/03-nouv/1782/jpg-0600/00000495.htm. See also en.wikipedia.org/wiki/Le_Sage's_theory_of_gravitation. In fact, the first to propose the idea of gravitation as a result of small particles pushing masses was Nicolas Fatio de Duillier in 1688. Cited on page 251.

188 G. 't Hooft, Dimensional reduction in quantum gravity, preprint at arxiv.org/abs/gr-qc/9310026. Many of the ideas of this paper become easier to understand and to argue when the strand model is used. Cited on page 255.

189 S. Carlip, Logarithmic corrections to black hole entropy from the Cardy formula, *Classical and Quantum Gravity* **17**, pp. 4175–4186, 2000, preprint at arxiv.org/abs/gr-qc/0005017. Cited on page 257.

190 On torsion, see the excellent review by R. T. Hammond, New fields in general relativity, *Contemporary Physics* **36**, pp. 103–114, 1995. Cited on page 264.

191 H. Kleinert & J. Zaanen, World nematic crystal model of gravity explaining the absence of torsion, *Physics Letters A* 324, pp. 361–365, 2004. Cited on page 264.

192 The analogy between the situation around line defects and general relativity is explained in Ekkehart Kröner, *Kontinuumstheorie der Versetzungen und Eigenspannungen*, Springer, 1958, These ideas have been taken up and pursued by J. D. Eshelby, B. A. Bilby, and many others after them. Cited on page 264.

193 Loop quantum gravity is a vast research field. The complete literature is available at arxiv.org/archive/gr-qc. Cited on page 264.

194 G. 't Hooft, Crystalline Gravity, *International Journal of Modern Physics A* 24, pp. 3243–3255, 2009, and also G. 't Hooft, A locally finite model of gravity, preprint at arxiv.org/abs/0804.0328. Cited on page 264.

195 van Raamsdonk, Comments on quantum gravity and entanglement, preprint at arxiv.org/abs/0907.2939. Cited on page 265.

196 C. H. Lineweaver & T. M. Davis, Misconceptions about the big bang, *Scientific American* pp. 36–45, March 2005. Cited on page 267.

197 Supernova Search Team Collaboration, A.G. Riess & al., Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astronomical Journal* 116, pp. 1009–1038, 1998, preprint at arxiv.org/abs/astro-ph/9805201. Cited on page 267.

198 Stephen Hawking & Roger Penrose, *The Nature of Space and Time*, Princeton University Press, 1996. Cited on page 268.

199 W. Fischler & L. Susskind, Holography and Cosmology, preprint at arxiv.org/abs/hep-th/9806039. Cited on page 269.

200 C. Balázs & I. Szapudi, Naturalness of the vacuum energy in holographic theories, preprint at arxiv.org/abs/hep-th/0603133. See also C. Bambi & F. R. Urban, Natural extension of the generalised uncertainty principle, preprint at arxiv.org/abs/0709.1965. The same point is made by D. A. Easson, P. H. Frampton & G. F. Smoot, Entropic accelerating universe, preprint at arxiv.org/abs/1002.4278. Cited on page 269.

201 For a review of recent cosmological data, see D.N. Spergel, R. Bean, O. Doré, M.R. Nolta, C.L. Bennett, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H.V. Peiris, L. Verde, C. Barnes, M. Halpern, R.S. Hill, A. Kogut, M. Limon, S.S. Meyer, N. Odegard, G.S. Tucker, J.L. Weiland, E. Wollack & E. L. Wright, Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology, preprint at arxiv.org/abs/astro-ph/0603449. Cited on pages 269 and 270.

202 There is a large body of literature that has explored a time-varying cosmological constant, especially in relation to holography. An example with many references is L. Xu, J. Lu & W. Li, Time variable cosmological constants from the age of the universe, preprint at arxiv.org/abs/0905.4773. Cited on page 270.

203 D. Wiltshire, Gravitational energy and cosmic acceleration, preprint at arxiv.org/abs/0712.3982; D. Wiltshire, Dark energy without dark energy, preprint at arxiv.org/abs/0712.3984. Cited on page 270.

204 The attribution to Voltaire could not be confirmed. Cited on page 273.

205 V. Crede & C. A. Meyer, The experimental status of glueballs, *Progress in Particle and Nuclear Physics* 63, pp. 74–116, 2009. Cited on page 283.

206 E. Klempt & A. Zaitsev, Glueballs, hybrids, multiquarks. Experimental facts versus QCD inspired concepts, *Physics Reports* **454**, 2007, preprint at arxiv.org/abs/0708.4016. Cited on pages 283 and 298.

207 R. V. Buniy & T. W. Kephart, A model of glueballs, preprint at arxiv.org/pdf/hep-ph/0209339; R. V. Buniy & T. W. Kephart, Universal energy spectrum of tight knots and links in physics, preprint at arxiv.org/pdf/hep-ph/0408025; R. V. Buniy & T. W. Kephart, Glueballs and the universal energy spectrum of tight knots and links, preprint at arxiv.org/pdf/hep-ph/0408027. See also J. P. Ralston, The Bohr atom of glueballs, preprint at arxiv.org/pdf/hep-ph/0301089. Cited on page 283.

208 See the one million dollar prize described at www.claymath.org/millennium/Yang-Mills_Theory. Cited on page 283.

209 A. J. Niemi, Are glueballs knotted closed strings?, pp. 127–129, in H. Suganuma, N. Ishii, M. Oka, H. Enyo, T. Hatsuda, T. Kunihiro & K. Yazaki editors, *Color confinement and hadrons in quantum chromodynamics*, World Scientific, 2003, preprint at arxiv.org/pdf/hep-th/0312133. See also Y. M. Cho, B. S. Park & P. M. Zhang, New interpretation of Skyrme theory, preprint at arxiv.org/pdf/hep-th/0404181; K. Kondo, A. Ono, A. Shibata, T. Shinohara & T. Murakami, Glueball mass from quantized knot solitons and gauge-invariant gluon mass, *Jornal of Physics A* **39**, pp. 13767–13782, 2006, preprint at arxiv.org/abs/hep-th/0604006. Cited on pages 283 and 298.

210 For a clear review on the topic and the planned experiments, see E. Fiorini, Measurement of neutrino mass in double beta decay, *Europhysics News* **38**, pp. 30–34, 2007, downloadable at www.europhysicsnews.org. Cited on page 286.

211 For example, see the detailed discussion of neutrino properties at pdg.web.cern.ch or, in print, in Ref. 218. Cited on page 286.

212 For a possible third approach, see A. F. Nicholson & D. C. Kennedy, Electroweak theory without Higgs bosons, *International Journal of Modern Physics A* **15**, pp. 1497–1519, 2000, preprint at arxiv.org/abs/hep-ph/9706471. Cited on page 287.

213 M. Veltman, The Higgs system, lecture slides at www.nikhef.nl/pub/theory/academiclectures/Higgs.pdf. See also his CERN Yellow Report 97-05, Reflections on the Higgs system, 1997, and the paper H. Veltman & M. Veltman, On the possibility of resonances in longitudinally polarized vector boson scattering, *Acta Physics Polonica B* **22**, pp. 669–695, 1991. Cited on page 287.

214 J. W. Moffat & V. T. Thot, A finite electroweak model without a Higgs particle, preprint at arxiv.org/abs/0812.1991. The ideas go back to D. Evens, J. W. Moffat, G. Kleppe & R. P. Woodard, Nonlocal regularizations of gauge theories, *Physical Review D* **43**, pp. 499–519, 1991. For more details on how to introduce non-locality while maintaining current conservation and unitarity, see G. Kleppe & R. P. Woodard, Non-local Yang-Mills, *Nuclear Physics B* **388**, pp. 81–112, 1992, preprint at arxiv.org/abs/hep-th/9203016. For a different approach that postulates no specific origin for the W and Z masses, see J. W. Moffat, Ultraviolet complete electroweak model without a Higgs particle, preprint at arxiv.org/abs/1006.1859. Cited on page 288.

215 H. B. Nielsen & P. Olesen, *A vortex line model for dual strings*, Nuclear Physics B, **61**, pp. 45-61 (1973). Cited on pages 290 and 323.

216 B. Andersson, G. Gustafson, G. Ingelman & T. Sjöstrand, Parton fragmentation and string dynamics, *Physics Reports* **97**, pp. 31–145, 1983. Cited on page 290.

217 C. B. Thorn, Subcritical string and large N QCD, preprint at arxiv.org/abs/0809.1085. Cited on page 290.

218 K. Nakamura & al., The Review of Particle Physics, *Journal of Physics G* **37**, p. 075021, 2010, or pdg.web.cern.ch. Cited on pages 293, 294, 295, 307, 313, 314, 315, 368, and 370.

219 A. J. Buchmann & E. M. Henley, Intrinsic quadrupole moment of the nucleon, *Physical Review C* **63**, p. 015202, 2000. Alfons Buchmann also predicts that the quadrupole moment of the other, strange $J = 1/2$ octet baryons is positive, and predicts a prolate structure for all of them (private communication). For the decuplet baryons, with $J = 3/2$, the quadrupole moment can often be measured spectroscopically, and is always negative. The four Δ baryons are thus predicted to have a negative intrinsic quadrupole moment and thus an oblate shape. This explained in A. J. Buchmann & E. M. Henley, Quadrupole moments of baryons, *Physical Review D* **65**, p. 073017, 2002. For recent updates, see A. J. Buchmann, Charge form factors and nucleon shape, pp. 110–125, in C. N. Papanicolas & Aron Bernstein editors, *Shape of Hadrons Workshop Conference*, Athens, Greece, 27-29 April 2006, AIP Conference Proceedings 904. Cited on pages 292 and 296.

220 A review on Regge trajectories and Chew-Frautschi plots is W. Drechsler, Das Regge-Pol-Modell, *Naturwissenschaften* **59**, pp. 325–336, 1972. See also the short lecture on courses.washington.edu/phys55x/Physics557_lec11.htm. Cited on page 293.

221 Kurt Gottfried & Victor F. Weisskopf, *Concepts of Particle Physics*, Clarendon Press, Oxford, 1984. Cited on page 293.

222 G. 't Hooft, G. Isidori, L. Maiani, A. D. Polosa & V. Riquer, A theory of scalar mesons, *Physics Letters B* **662**, pp. 424–430, 2008, preprint at arxiv.org/abs/0801.2288. Cited on page 298.

223 J. Viro & O. Viro, Configurations of skew lines, *Leningrad Mathematical Journal* **1**, pp. 1027–1050, 1990, and updated preprint at arxiv.org/abs/math.GT/0611374. Cited on page 300.

224 W. Thomson, On vortex motion, *Transactions of the Royal Society in Edinburgh* pp. 217–260, 1868. This famous paper stimulated much work on knot theory. Cited on page 300.

225 H. Jehle, Flux quantization and particle physics, *Physical Review D* **6**, pp. 441–457, 1972, and H. Jehle, Flux quantization and fractional charge of quarks, *Physical Review D* **6**, pp. 2147–2177, 1975. Cited on page 301.

226 T. R. Mongan, Preons from holography, preprint at arxiv.org/abs/0801.3670. Cited on page 301.

227 Jacob's rings are shown, for example, in the animation on www.prestidigitascience.fr/index.php?page=anneaux-de-jacob. They are already published in the book by Tom Tit, *La science amusante*, 1870, and the images were reprinted the popular science books by Edi Lammers, and, almost a century later on, even in the mathematics column and in one of the books by Martin Gardner. Cited on page 302.

228 P. Pieranski, S. Przybyl & A. Stasiak, Tight open knots, *European Physical Journal E* **6**, pp. 123–128, 2001, preprint at arxiv.org/abs/physics/0103016. Cited on pages 306, 307, and 373.

229 R. Boughezal, J. B. Tausk & J. J. van der Bij, Three-loop electroweak corrections to the W-boson mass and $\sin^2\theta_{\mathrm{eff}}$ in the large Higgs mass limit, *Nuclear Physics B* **725**, pp. 3–14, 2005, preprint at arxiv.org/abs/hep-ph/0504092. Cited on page 307.

230 There is an interesting exploration behind this analogy. The first question is to discover why this analogy between a non-dissipative system – a free quantum particle moving in vacuum – and a dissipative system – a macroscopic body drawn through a viscous liquid, say honey – is possible at all. (A careful distinction between the cases with spin 0, spin 1 and spin 1/2

Challenge 166 e

are necessary.) The second question is the exploration of the motion of bodies of general shape in viscous fluids at low Reynolds numbers and under constant force. For the best overview of this question, see the beautiful article by O. Gonzalez, A. B. A. Graf & J. H. Maddocks, Dynamics of a rigid body in a Stokes fluid, *Journal of Fluid Mechanics* **519**, pp. 133–160, 2004. Cited on page 308.

231 E. Rawdon & M. Fisher, private communication. Cited on pages 308 and 309.

232 See H. Fritzsch, A.D. Özer, A scaling law for quark masses, preprint at arxiv.org/abs/hep-ph/0407308. Cited on page 309.

233 K. A. Meissner & H. Nicolai, Neutrinos, axions and conformal symmetry, preprint at arxiv.org/abs/0803.2814. Cited on pages 310 and 311.

234 M. Shaposhnikov, Is there a new physics between electroweak and Planck scale?, preprint at arxiv.org/abs/0708.3550. Cited on page 311.

235 Y. Diao, C. Ernst, A. Por & U. Ziegler, The roplength of knots are almost linear in terms of their crossing numbers, preprint at arxiv.org/abs/0912.3282. Cited on page 311.

236 H. Fritzsch & Z. -Z. Xing, Lepton mass hierarchy and neutrino mixing, preprint at arxiv.org/abs/hep-ph/0601104 Cited on page 315.

237 The effects of neutrino mixing, i.e., neutrino oscillations, were measured in numerous experiments from the 1960s onwards; most important were the experiments at Super-Kamiokande in Japan and at the Sudbury Neutrino Observatory in Canada. See Ref. 218. Cited on page 315.

238 M. Fukugita & T. Yanagida, Baryogenesis without grand unification, *Physics Letters B* **174**, pp. 45–47, 1986. Cited on page 316.

239 J. M. Cline, Baryogenesis, preprint at axiv.org/abs/0609145. Cited on page 316.

240 Several claims that the coupling constants changed with the age of the universe have been refuted by all subsequent measurements. Cited on page 319.

241 That tight tangles correlate with random tangles was first shown by V. Katritch, J. Bednar, D. Michoud, R.G. Sharein, J. Dubochet & A. Stasiak, Geometry and physics of knots, *Nature* **384**, pp. 142–145, 1996. It was confirmed by E.J. Janse van Rensburg, E. Orlandini, D.W. Sumners, M.C. Tesi & S.G. Whittington, The writhe of knots in the cubic lattice, *Journal of Knot Theory and its Ramifications* **6**, pp. 31–44, 1997. Cited on page 322.

242 The quasi-quantization of writhe was discovered by V. Katritch, J. Bednar, D. Michoud, R. G. Scharein, J. Dubochet & A. Stasiak, Geometry and physics of knots, *Nature* **384**, pp. 142–145, 1996. See also P. Pieranski, In search of ideal knots, pp. 20–41, and A. Stasiak, J. Dubochet, V. Katritch & P. Pieranski, Ideal knots and their relation to the physics of knots, pp. 1–19, both found in A. Stasiak, V. Katritch & L. H. Kauffman editors, *Ideal Knots*, World Scientific, 1998. Most pedagogic is P. Pieranski & S. Przybyl, Quasi-quantization of writhe in ideal knots, *European Physical Journal E* **6**, pp. 117–121, 2001, preprint at arxiv.org/abs/physics/0102067. See also C. Cerf & A. Stasiak, Linear behavior of the writhe versus the number of crossings in rational knots and links, pp. 111–126, in M. I. Monastyrsky editor, *Topology in Molecular Biology*, 2007. The quasi-quantization, and in particular the lack of precise quantization, was confirmed in 2010 by T. Ashton, J. Cantarella, M. Piatek & E. Rawdon private communication. Cited on page 322.

243 D. Hilbert, Über das Unendliche, *Mathematische Annalen* **95**, pp. 161–190, 1925. Cited on page 330.

244 The *Book of Twenty-four Philosophers*, *c.* 1200, is attributed to Hermes Trismegistos, but was actually written in the middle ages. The text can be found in F. Hudry, ed., *Liber viginti quattuor philosophorum*, Turnholt, 1997, in the series *Corpus Christianorum, Continuatio Mediaevalis*, CXLIII a, tome III, part 1, of the Hermes Latinus edition project headed by P. Lucentini. There is a Spinozian cheat in the quote: instead of 'nature', the original says 'god'. The reason why this substitution is applicable is given earlier on. Cited on page 335.

Page 349

245 As a disappointing example, see Gilles Deleuze, *Le Pli – Leibniz et le baroque*, Les Editions de Minuit, 1988. In this unintelligible, completely crazy book, the author pretends to investigate the implications of the idea that the *fold* (in French 'le pli') is the basic entity of matter and 'soul'. Cited on page 336.

246 René Descartes, *Discours de la méthode*, 1637. He used and discussed the sentence again in his *Méditations métaphysiques* 1641, and in his *Les principes de la philosophie* 1644. Both books influenced many thinkers in the subsequent centuries. Cited on page 341.

247 D. D. Kelly, Sleep and dreaming, in *Principles of Neural Science*, Elsevier, New York, 1991. The paper summarises experiments made on numerous humans and shows that even during dreams, people's estimate of time duration corresponds to that measured by watches. Cited on page 341.

248 Astrid Lindgren said this in 1977, in her speech at the fiftieth anniversary of Oetinger Verlag, her German publisher. The German original is: 'Alles was an Großem in der Welt geschah, vollzog sich zuerst in der Phantasie eines Menschen, und wie die Welt von morgen aussehen wird, hängt in großem Maß von der Einbildungskraft jener ab, die gerade jetzt lesen lernen.' The statement is found in Astrid Lindgren, Deshalb brauchen Kinder Bücher, *Oetinger Almanach* Nr. 15, p. 14, 1977. Cited on page 343.

CREDITS

Acknowledgments

In the first half of the text, stimulating discussions in the 1990s with Luca Bombelli helped in structuring the chapter on the contradictions between general relativity and quantum theory, as well as the chapter on the difference between vacuum and matter. In the years up to 2005, stimulating discussions with Saverio Pascazio, Corrado Massa and especially Steven Carlip helped shaping the chapter on limit values.

The second half of the text, on the strand model, owes much to Louis Kauffman. The ideas found in his books and in his papers were the basis for this text long before we met and exchanged mails. In particular, his ideas on knots, on quantum theory, on measurement, on particle physics, on set theory and on foundational issues were a constant inspiration. His papers – available on www2.math.uic.edu/~kauffman – and his books are all worth reading; among them is the fascinating paper *Knot logic* and the wonderful text *Knots and Physics*, World Scientific, 1991. These publications convinced me that strands and knots are a promising direction in the search for a final theory. The breadth of Lou's knowledge and interests, the depth of his passion, and his warm humanity are exemplary.

I thank Eric Rawdon for his ropelength calculations. I also thank Claus Ernst, Andrzej Stasiak, Ralf Metzler, and Jason Cantarella for their input and the fruitful discussions we had. Also Saverio Pascazio, Roland Netz, Christian Weitzel, Hans Aschauer, Stephan Schiller, Gerrit Bauer, Richard Hoffmann, Axel Schenzle, Reinhard Winterhoff, Alden Mead, Franca Jones-Clerici and Damoon Saghian provided valuable help. My parents, Isabella and Peter Schiller, strongly supported the project. I thank my mathematics and physics teacher in secondary school, Helmut Wunderling, for the fire he has nurtured inside me. I also thank the lawmakers and the taxpayers in Germany, who, in contrast to most other countries in the world, allow residents to use the local university libraries.

The typesetting and the final pdf file were improved with help of Ulrich Dirr, Johannes Kuester and Michael Zedler. Kevin Bicknell and Donald Arseneau helped with LaTeX macros. The numerous other experts who helped with the intricacies of LaTeX are listed in the acknowledgements of the previous volumes.

Since May 2007, the electronic edition and distribution of the Motion Mountain text is generously supported by the Klaus Tschira Foundation.

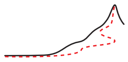
But above all, it was the continuous support of my wife Britta that made this volume possible.

Film credits

The animations of the belt trick on page 170 are copyright and courtesy of Greg Egan; they can be found on his website www.gregegan.net/APPLETS/21/21.html. I am also grateful to Greg Egan for expanding his applet so as to show both possible options of the belt trick. Greg Egan's animations were included into the pdf file with the help of a copy of the iShowU software sponsored by Neil Clayton at www.shinywhitebox.com.

Image credits

The mountain photograph on the front cover is courtesy and copyright by Dave Thompson and found on his website www.daveontrek.co.uk. The photograph of the Ultra Deep Field project on page 16 is courtesy of NASA. The drawing by Maurits Escher on page 55 is copyright by the M.C. Escher Heirs, c/o Cordon Art, Baarn, The Netherlands, who kindly gave permission for its use. The drawing by Peter Battey-Pratt and Thomas Racey of the belt trick on page 199, taken from Ref. 174, is courtesy and copyright by Springer Verlag. The images of the tight open knots on page 306 are courtesy and copyright by Piotr Pieranski, and found in Ref. 228. The graph of the running coupling constants on page 318 is courtesy and copyright by Wim de Boer and taken from his home page at www-ekp.physik.uni-karlsruhe.de/~deboer. The photograph of Cerro Torre on page 338 is copyright and courtesy of Davide Brighenti, and found on Wikimedia Commons; the photograph of the green hill on the same page is copyright and courtesy of Myriam70, and found on her site www.flickr.com/photos/myriam70. The photograph on the back cover, of a basilisk running over water, is courtesy and copyright by the Belgian group TERRA vzw and found on their website www.terra.vzw.org. All drawings are copyright by Christoph Schiller.



NAME INDEX

Page numbers in *italic* typeface refer to pages where the person is presented in more detail.

SUBJECT INDEX

Page numbers in *italic* typeface refer to pages where the keyword is defined or presented in detail. The subject index thus acts as a glossary.

MOTION MOUNTAIN

The Adventure of Physics – Vol. VI

A Speculation On Unification

Which problems in physics were unsolved in the year 2000?
What might be their solution?
Is ‘empty space’ really empty?
At what distance between two points does it become
impossible to find room for a third one in between?
What is the most fantastic voyage possible?
Why do change and motion exist?

Answering these and other questions on motion, this book gives an entertaining and mind-twisting introduction into modern research on the unification of physics: it presents the *strand model*. Based on a simple principle, strands reproduce quantum theory, the standard model of particle physics and general relativity. While leaving no room for alternative theories, strand agree with all experimental data and make surprising predictions.

Christoph Schiller, PhD Université Libre de Bruxelles, is a physicist and physics popularizer. He wrote this book for students, teachers and anybody interested in modern physics.