Christoph Schiller

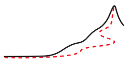
MOTION MOUNTAIN

THE ADVENTURE OF PHYSICS – VOL. VI

THE STRAND MODEL –

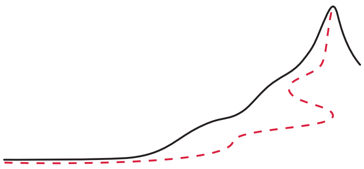
A SPECULATION ON UNIFICATION

www.motionmountain.net



Christoph Schiller

MOTION MOUNTAIN



The Adventure of Physics
Volume VI

The Strand Model –
A Speculation on Unification

Edition 30, available as free pdf
with films at www.motionmountain.net

Editio trigesima.

Thirtieth edition.

To Britta, Esther and Justus Aaron

τῷ ἐμοὶ δαὶμονι

Die Menschen stärken, die Sachen klären.

PREFACE

This book is for anybody who is intensely curious about motion. Why and how do things, people, trees, stars, images or 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 aim to describe *all* motion – everyday, quantum and relativistic – implies a large project. This project can be structured using the diagram shown in Figure 1, the so-called *Bronshtein cube*. The previous volumes have covered all points in the cube – all domains of motion – except the highest one. This remaining point contains the *final* and *unified* description of all motion. The present volume briefly summarizes the history of this old quest and then presents an intriguing, though speculative solution to the riddle.

The search for the final, unified description of motion is a story of many surprises. First, twentieth-century research has shown that there is a smallest measurable distance in nature, the Planck length. Then it appeared that matter cannot be distinguished from empty space at those small distances. A last surprise dates from this century: particles and space appear to be made of *strands*, instead of little spheres or points. The present text explains how to reach these surprising 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, including the origin of colours and of the parameters of the standard model, appear to be answerable.

The strand model, as presented in this text, is an unexpected result from a threefold aim that the author has 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. While the previous volumes introduced the *established* parts of physics, this volume presents, in the same captivating and playful way, a *speculation* about unification. Nothing in this volume is established knowledge – yet. The text is the original presentation of the topic. The aim for maximum simplicity has been central in deducing this speculation.

The search for a final theory is one of the great adventures of life: it leads to the limits of thought. The search overthrows several of our thinking habits about nature. This can produce fear, but by overcoming it we gain strength and serenity. Changing thinking habits requires courage, but it produces intense and beautiful emotions. Enjoy them.

Munich, summer 2017.

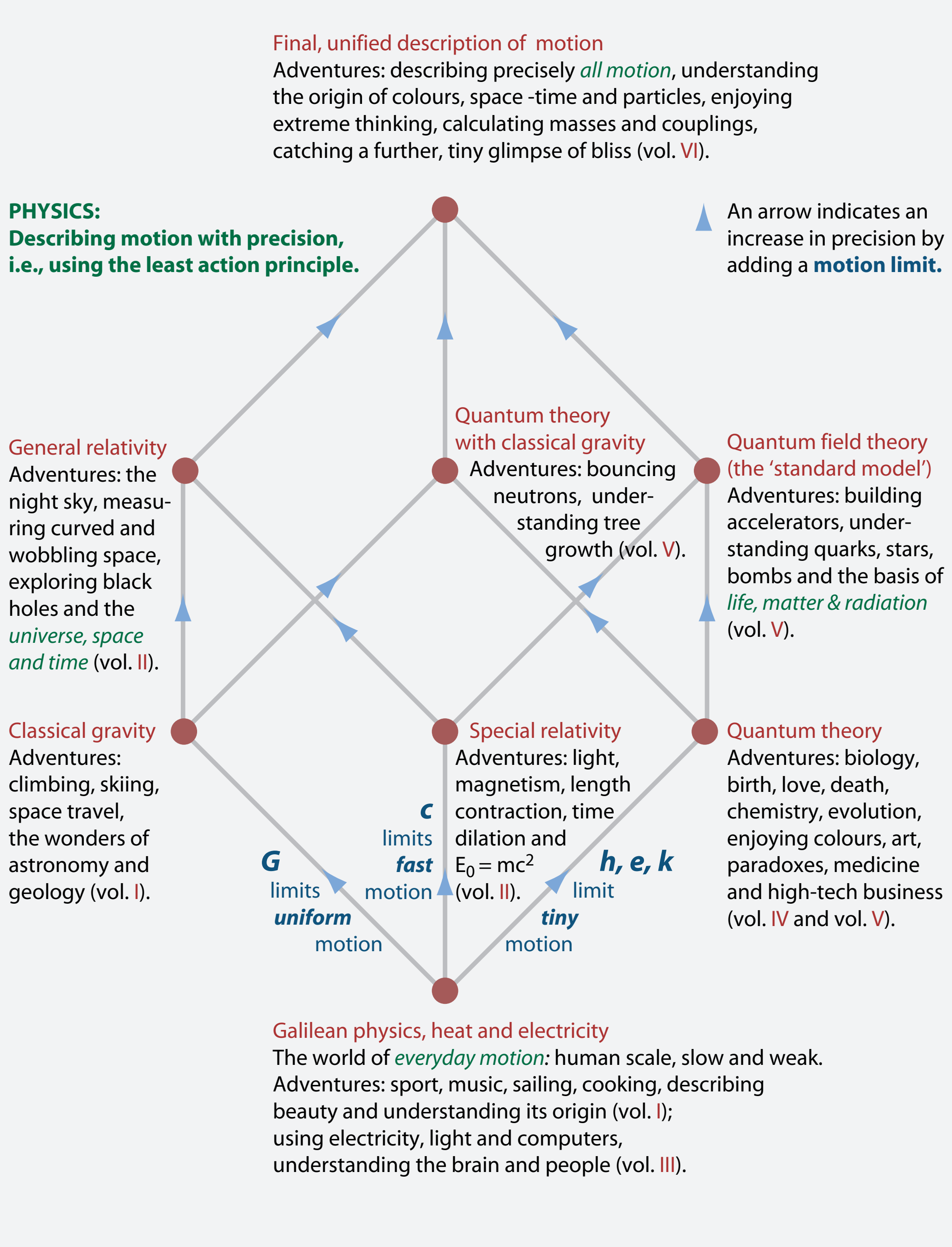


FIGURE 1 A complete map of physics, the science of motion, as first proposed by Matvei Bronshtein (b. 1907 Vinnytsia, d. 1938 Leningrad). The map is of central importance in the present volume. The Bronshtein cube starts at the bottom with everyday motion, and shows the connections to the fields of modern physics. Each connection increases the precision of the description and is due to a limit to motion that is taken into account. The limits are given for uniform motion by the gravitational constant G, for fast motion by the speed of light c, and for tiny motion by the Planck constant h, the elementary charge e and the Boltzmann constant k.

USING THIS BOOK

To get a *quick overview*, read the first chapter and continue with the summary sections only. There are summaries at the end of each chapter. In addition, throughout the text,

▷ Important ideas are marked with a triangle.

Marginal notes refer to bibliographic references, to other pages or to challenge solutions. In the colour edition, such notes and also the pointers to footnotes and to other websites are typeset in green. In the free pdf edition of this book, available at www.motionmountain.net, all green pointers and links are clickable. The pdf edition also contains all films; they can be watched directly in Adobe Reader. Over time, links on the internet tend to disappear. Most links can be recovered via www.archive.org, which keeps a copy of old internet pages.

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

A *paper edition* of this book is available, either in colour or in black and white, from www.amazon.com or www.createspace.com. So is a Kindle edition.

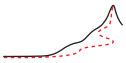
FEEDBACK

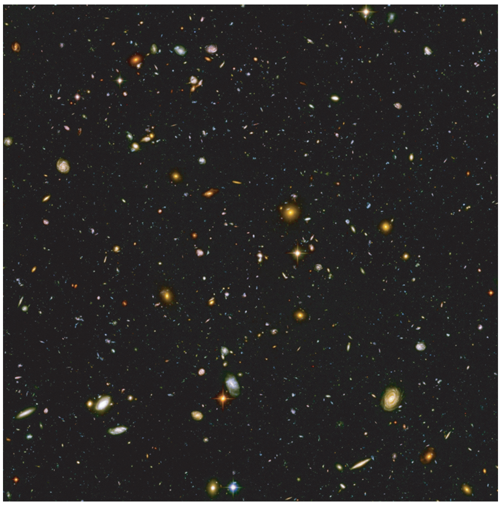
Receiving an email from you at fb@motionmountain.net, either on how to improve the text or on a solution for one of the prize challenges mentioned on www.motionmountain.net/prizes.html, would be delightful. All feedback will be used to improve the next edition. For a particularly useful contribution you will be mentioned – if you want – in the acknowledgements, receive a reward, or both.

SUPPORT

Your donation to the minuscule, charitable, tax-exempt non-profit organisation that produces, translates and publishes this book series is welcome. For details, see the web page www.motionmountain.net/donation.html. The German tax office checks the proper use of your donation. If you want, your name will be included in the sponsor list. Thank you in advance for your help, on behalf of all readers across the world. And now, enjoy the reading.

CONTENTS





THE STRAND MODEL – 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 gauge interactions appear naturally,
that colours are due to strand twisting,
and that a complete description of motion is possible.

CHAPTER 1

FROM MILLENNIUM PHYSICS TO UNIFICATION

Look at what happens around us. A child that smiles, a nightingale that sings, a lily that opens: all move. Every shadow, even an immobile one, is due to moving light. Every mountain is kept in place by moving electrons. Every star owes its formation and its shine to motion of matter and radiation. Also the darkness of the night sky[**] is due to motion: it results from the expansion of space. Finally, human creativity is due to the motion of molecules, ions and electrons in the brain. Is there a common language for these and all other observations of nature?

Is there a *unified* and *precise* way to describe all motion? How? 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, researchers collected a huge number of precise observations about motion. We now know how electric signals move in the brain, how insects fly, why colours vary, 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 both useful and fascinating.

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

1. 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
2. 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
3. In nature, there is an invariant highest momentum flow, the Planck force $c^4/4G$. This invariant maximum implies *general relativity*, as we will recall below. Among others, Page 30

[**] 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).

general relativity implies that things fall and that empty space curves and moves, as is observed. Ref. 2

4. 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

5. In nature, there is a non-zero, invariant smallest change value, 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

6. 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. In addition, the year 2012 finally brought the discovery of the Higgs boson, which was already predicted in 1964. 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, why colours vary, and how the atoms in our bodies came to be. Ref. 4

These six statements, the *millennium description of physics*, describe everything known about motion in the year 2000. (Actually, 2012 is a more precise, though less striking date.) These statements describe the motion of people, animals, plants, objects, light, radiation, stars, empty space and the universe. The six statements describe motion so precisely that even today there is *no* difference between calculation and observation, between theory and practice. This is an almost incredible result, the summary of the efforts of tens of thousands of researchers during the past centuries.

However, a small set of observations does not yet follow from the six statements. A famous example is the origin of colours. In nature, colours are consequences of the so-called *fine structure constant*, a mysterious constant of nature, abbreviated α, whose value is measured to be $\alpha = 1/137.035\,999\,139(31)$. If α had another value, all colours would differ. And why are there *three* gauge interactions, *twelve* elementary fermions, *thirteen* elementary bosons and *three* dimensions? What is the origin of particle masses? Why is the standard model, the sixth statement above, so complicated? How is it related to the five preceding statements? Ref. 5

A further unexplained observation is the nature of dark matter found around galaxies. We do not know yet what it is. Another unexplained process is the way thinking forms in our brain. We do not know yet in detail how thinking follows from the above six statements, though we do know that thinking is not in contrast with them. For this reason, we will not explore the issue in the following. In the case of dark matter this is not so clear: dark matter could even be in contrast with the millennium description of motion.

Finally, why is there motion anyway? In short, even though the millennium description of physics is precise and successful, it is not complete. The list of all those fundamental issues about motion that are *unexplained* since the year 2000 make up only a short table. We call them the *millennium issues*.

TABLE 1 The millennium list: *everything* the standard model 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 since the year 2000
Local quantities unexplained by the standard model: particle properties	
$\alpha = 1/137.036(1)$	the low energy value of the electromagnetic coupling or fine structure 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
Concepts unexplained by the standard model	
$c, \hbar, k$	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
Ψ	the origin and nature of wave functions
$S(n)$	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, of the vanishing of magnetic charge, and of minimal coupling
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
Renorm. group	the origin of renormalization properties
$\delta W = 0$	the origin of 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 unexplained by general relativity and cosmology	
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)
$\rho_\mathrm{de} = \Lambda c^4/(8\pi G)$ $\approx 0.5\,\mathrm{nJ/m^3}$	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
ρ_dm	the density and nature of dark matter

TABLE 1 (Continued) The millennium list: *everything* the standard model and general relativity *cannot* explain; also the *only* experimental data available to test the final, unified description of motion.

Observable	Property unexplained since the year 2000
$f_0(1, ..., c.\ 10^{90})$	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
Concepts unexplained by general relativity and cosmology	
c, G	the origin of the invariant Planck units of general relativity
$\mathrm{R} \times \mathrm{S}^3$	the observed topology of the universe
$G^{\mu\nu}$	the origin and nature of curvature, the metric and horizons
$\delta W = 0$	the origin of the least action principle in general relativity
$W = \int L_{\mathrm{GR}} \mathrm{d}t$	the origin of the Lagrangian of general relativity

The millennium list contains *everything* that particle physics and general relativity *cannot* explain. In other words, the list contains *every issue* that was unexplained in the domain of fundamental motion in the year 2000. The list is short, but it is not empty. Every line in the millennium list asks for an explanation. The quest for unification – and the topic of this text – is the quest for these explanations. We can thus say that a *final theory of motion* is a theory that eliminates the millennium list of open issues.

AGAINST A FINAL THEORY

We know that a final theory exists: it is the theory that describes how to calculate the fine structure constant $\alpha = 1/137.036(1)$. The theory does the same for about two dozen other constants, but α is the most famous one. In other terms, the final theory is the theory that explains all colours found in nature.

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 claimed 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 just 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. Page 167
- Some state that it is not clear whether a final theory exists at all. We all know from experience that this is wrong. The reason is simple: *We are able to talk about everything.* In other words, all of us already have a 'theory of everything', or a final theory of

nature. Also a physical theory is a way to talk about nature, and for the final theory we only have to search for those concepts that enable us to talk about all of motion *with full precision*. Because we are just looking for a way to talk, we know that the final theory exists. And searching for it is fascinating and exciting, as everybody busy with this adventure will confirm.

— 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

— 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 child education problems required – though this might help. Ref. 7 Vol. IV, page 141

— Some people claim that searching for a final theory is a sign of foolishness or a sin of pride. Such defeatist or envious comments should simply be ignored. After all, the quest is just the search for the solution to a riddle.

— 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. Ref. 8

— 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 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. Ref. 9

— 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.

These arguments show us that 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. (We get rid of all baggage in the first six chapters of this volume.) 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. The search for unification is a pastime. Any riddle is best approached with the

Ref. 10 light-heartedness of playing. Life is short: we should play whenever we can.

WHAT WENT WRONG IN THE PAST

The twentieth century was the golden age of physics. Scholars searching for the final theory explored candidates such as grand unified theories, supersymmetry and numerous other options. These candidates will be discussed later on; all were falsified by 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? Vol. V, page 267

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.’
- ‘Unification requires finding higher symmetries.’ (Werner Heisenberg)
- ‘Unification requires generalizing electroweak mixing to include the strong interaction.’ (Abdus Salam)
- ‘Unification requires extending the standard model of particle physics with supersymmetry.’ (Steven Weinberg)
- ‘Unification requires axiomatization.’ (David Hilbert)
- ‘Unification requires searching for beauty.’ (Paul Dirac)
- ‘Unification requires new quantum evolution equations.’ (Werner Heisenberg)
- ‘Unification requires new field equations of gravitation.’ (Albert Einstein)
- ‘Unification requires more dimensions of space.’ (Theodor Kaluza)
- ‘Unification requires topology change.’ (John Wheeler)
- ‘Unification is independent of Planck’s natural units.’
- ‘Unification requires using complicated mathematics and solving huge conceptual difficulties.’ (Edward Witten)
- ‘Unification is only for a selected few.’
- ‘Unification is extremely useful, important and valuable.’

All these beliefs appeared in the same way: first, some famous scholar – in fact, many more than those mentioned – explained the idea that guided his discovery; then, he and most other researchers started to believe the guiding idea more than the discovery itself. The most explored belief were those propagated by Salam and Weinberg: they – unknowingly – set thousands of researchers on the wrong path for dozens of years. The most detrimental has been the belief that unification is complicated and difficult: it kept the smartest physicists from producing progress. In fact, all the mentioned beliefs can be seen as special cases of the first one. And like the first belief, they are all, as we will discover in the following, wrong.

AN ENCOURAGING ARGUMENT

Page 8 The Bronshtein cube in Figure 1 shows that physics started from the description of motion in everyday life. At the next level of precision, physics introduced the observed *limits* to motion and added the description of powerful, i.e., as uniform as possible motion (classical gravity), as fast as possible motion (special relativity), and as tiny as possible motion (quantum theory). At the following level of precision, physics achieved all possible combinations of two of these motion types, by taking care of two motion limits at the same time: fast and uniform motion (general relativity), fast and tiny motion (quantum field theory), and tiny and uniform motion (quantum theory with gravity). The only domain left over is the domain where motion is fast, tiny and as uniform as possible at the same time. When this last domain is reached, the precise description of all motion is completed.

But Figure 1 suggests even stronger statements. First of all, no domain of motion is left: the figure covers *all* motion. Secondly, the final description appears when general relativity, quantum field theory and quantum theory with gravity are combined. In other words, the final theory appears when relativity and quantum theory and interactions are all described together. But a third conclusion is especially important. Each of these three fields can be deduced from the unified final theory by eliminating a limitation: either that of tiny motion, that of straight motion, or that of fast motion. In other words:

▷ General relativity follows from the final theory by eliminating the quantum of action $\hbar$, i.e., taking the limit $\hbar \rightarrow 0$.
▷ Quantum field theory, including quantum electrodynamics, follows from the final theory by eliminating G, i.e., taking the limit $G \rightarrow 0$.
▷ Quantum theory with gravity follows from the final theory by eliminating the speed limit c, i.e., taking the limit $1/c \rightarrow 0$.

Speaking even more bluntly, and against a common conviction of researchers in the field, the figure suggests: *The standard model follows from the final theory by eliminating gravity.* These connections eliminate many candidates for the unified final theory that were proposed in the research literature in the twentieth and twenty-first century. But more importantly, the connections leave open a range of possibilities – and interestingly enough, this range is very narrow.

The figure allows stronger statements still. Progress towards the final theory is achieved by taking limitations to motion into account. Whatever path we take from everyday physics to the final theory, we must take into consideration all limits to motion. The order can differ, but *all* limits have to be taken into account. Now, if any intermediate steps – due to additional motion limitations – between quantum field theory and the final theory existed in the *upper* part of the figure, corresponding steps would have to appear also in the *lower* part of the figure, between everyday physics and classical gravity. In the same way, if any intermediate limits between general relativity and the final theory really existed, these limits would also have to appear between everyday motion and quantum theory.

Experiments show that *no* intermediate steps or limits exist between everyday motion and the next level of precision. Using the top-down symmetry of Figure 1, this implies:

▷ Intermediate steps or theories do not exist before the final theory.

This is a strong statement. In the foundations of motion, apart from the final theory, no further theory is missing. For example, the figure implies that there is no separate theory of relativistic quantum gravity or no doubly special relativity.

In particular, Figure 1 implies that, conceptually, we are already *close* to the final theory. The figure suggests that there is no need for overly elaborate hypotheses or concepts to reach the final theory.

▷ We just have to add G to the standard model or $\hbar$ and e to general relativity.

In short, the final, unified theory of motion cannot be far.

SUMMARY: 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 achieve this, we need to find a final theory that solves and explains each open issue given in the *millennium list*. This is our starting point.

We proceed in steps. We first simplify quantum theory and gravitation as much as possible, we explore what happens when the two are combined, and we 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* for the proposed model.

In short, three lists structure our quest for a final theory: the millennium list of open issues, the list of requirements for the final theory, 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 *extended*, without ends, one-dimensional and fluctuating: 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*.

Page 147 At this point, after the first half of our adventure, we obtain a detailed *requirement list* for the final theory. This list allows us to proceed rapidly to our goal, without being led astray.

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 a unique way. 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*, and all their observed properties. Other elementary particles do not exist. We thus recover the standard model of elementary particles.
10. We discover that the fundamental principle allows us to solve all the issues in the millennium list, and that all properties deduced from strands agree with experiment. In particular, the strand model allows us to calculate the fine structure constant and Page 393 the other gauge coupling strengths. An extensive *list of testable predictions* can be given. These predictions will all be tested – by experiment or by calculation – in the coming years.
11. We discover that motion is due to crossing switches of strands. Motion is an inescapable consequence of observation: motion is an experience that we make because we are, like every observer, 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.* Everything presented in the following agrees with experiment. Nevertheless, with almost every sentence you will find at least one physicist or philosopher who disagrees. That makes the adventure even more fascinating.

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

---

* ‘It is almost impossible to carry the torch of truth through a crowd without scorching somebody’s beard.’ Georg Christoph Lichtenberg (b. 1742 Ober-Ramstadt, d. 1799 Göttingen) was a famous physicist and essayist.

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.

Everyday, or Galilean, physics in one statement

Everyday motion is described by Galilean physics. It consists of only one statement:

▷ Motion minimizes change.

In nature, *change* is measured by physical action W. More precisely, change is measured by the time-averaged difference between kinetic energy T and potential energy U. In other words, motion obeys the so-called *least action principle*, written as

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

This statement determines the effort we need to move or throw stones, and explains why cars need petrol and people need food. In other terms, *nature is as lazy as possible*. Or:

▷ Nature is maximally efficient.

Vol. I, page 28 The efficiency or laziness of nature implies that motion is conserved, relative and predictable. In fact, the laziness of motion and nature is valid throughout modern physics, for all observations, 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: Ref. 11

▷ There is a maximum energy speed value c in nature.

For all physical systems and all observers, the local energy speed v is limited by the speed of light c:

$$v \leqslant c = 3.0 \cdot 10^8 \text{ m/s} . \tag{2}$$

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 speed limit is valid both for composite systems and for elementary particles. No exception has ever been found. (Try it.) Challenge 1 e

The energy speed limit is an *invariant*: the local 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 99

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 so-called *(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 26

---

* 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.

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 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: Ref. 12

▷ There is a minimum action value $\hbar$ in nature.

Vol. IV, page 14 For all physical systems and all observers, the action W obeys

$$W \geqslant \hbar = 1.1 \cdot 10^{-34}\,\mathrm{Js} . \tag{5}$$

The Planck constant $\hbar$ is the smallest observable action value, and the smallest observable change of angular momentum. The action limit is valid for all systems, thus both for composite and elementary systems. No exception has ever been found. (Try it.) The principle contains all of quantum theory. We call it the *principle of non-zero action*, in order to avoid confusion with the principle of least action. Challenge 3 e

The non-zero action limit $\hbar$ is an *invariant*: it is valid with the same numerical value for *all* observers. Again, any such 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 where the process involves 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 $\hbar$. Since physical action measures the change in the state of a physical system, 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

---

* 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. Challenge 4 e

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, 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. 13 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. Whenever we try to overcome the smallest action value, the experimental outcome is probabilistic. The minimum action value also implies that in quantum theory, the three 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*. Finally, the non-zero action limit is also the foundation of Einstein–Brillouin–Keller quantization. Ref. 14

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 usual 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 bounds for speed and for action for any physical system, $v \leqslant c$ and $W \leqslant pd \leqslant mcd$, when combined with the quantum of action, imply a limit on the displacement d of a system between any two observations:

$$d \geqslant \frac{\hbar}{mc} . \tag{6}$$

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 this quantum displacement limit also applies to elementary systems, it also applies to the *size* of a *composite* system. However, for the same reason, this size limit is *not* valid for the sizes of elementary particles. Challenge 5 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 23

$$\Delta d \, \Delta p \geqslant \frac{\hbar}{2} . \tag{7}$$

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 value k in nature.

Written symbolically,

$$S \geqslant k = 1.3 \cdot 10^{-23}\,\mathrm{J/K}\,. \qquad (8)$$

Challenge 6 e

Ref. 15

The entropy S is limited by the Boltzmann constant k. No exception has ever been found. (Try it.) 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.

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:

$$\Delta\frac{1}{T}\,\Delta U \geqslant \frac{k}{2}\,. \qquad (9)$$

Ref. 16

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

GENERAL RELATIVITY IN ONE STATEMENT

This text can be enjoyed most when a compact and unconventional description of general relativity is used; it is presented in the following. However, the conclusions do not depend on this description; the results are also valid if the usual approach to general relativity is used; this will be shown later on.

Page 54

The most compact description summarizes the step from universal gravity to general relativity in a single statement on motion:

▷ There are maximum force and power values in nature.

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

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

Challenge 7 e

No exception has ever been found. (Try it.) These limit statements contain both the speed of light c and the gravitational constant G; they thus qualify as statements about relativistic gravitation. Before we deduce general relativity, let us explore these limits.

The numerical values of the limits are huge. The maximum power corresponds to converting 50 solar masses into massless radiation within 1 millisecond. And applying the maximum force value along a distance l costs as much energy as a black hole of diameter l.

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. This is a simple consequence of the continuity equation. Therefore, every discussion of maximum force implies a clarification of the underlying surface.

Ref. 17

Vol. I, page 228

Both the force and the power limits state that the flow of momentum or of energy through any *physical surface* – a surface to which an observed can be attached at every one of its points – of any size, for any observer, in any coordinate system, never exceeds the limit value. In particular:

▷ The force limit is only realized *at horizons*. The power limit is only realized *with 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 109

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.

An object of mass m that has the size of its own Schwarzschild radius $2Gm/c^2$ is called a *black hole*, because according to general relativity, no signals and no light from inside the Schwarzschild radius can reach the outside world. In this text, black holes are usually non-rotating and usually uncharged; in this case, the terms 'black hole' and 'Schwarzschild black hole' are synonymous.

Ref. 18

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. (This is the so-called *hoop conjecture*.) In fact, the mass–energy concentration limit can easily be transformed algebraically into the force limit: they are equivalent.

Challenge 8 e

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 9 e

But is it possible to *imagine* a system that exceeds the force limit? An extensive discussion shows that this is impossible. For example, the force limit cannot be overcome with Lorentz boosts. We 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 relativity, 3-force cannot be increased beyond all bounds using boosts. In all reference frames, the measured 3-force can never exceed the

Vol. II, page 106

Vol. II, page 82

proper force, i.e., the 3-force value measured in the comoving frame.

Also changing to an accelerated frame does not help to overcome the force limit, because for high accelerations a, horizons appear at distance c^2/a, and a mass m has a minimum diameter given by $l \geqslant 4Gm/c^2$.

In fact, the force and power limits cannot be exceeded in any thought experiment, Vol. II, page 106 as long as the sizes of observers or of test masses are taken into account. All apparent exceptions or paradoxes assume the existence of point particles or point-like observers; these, however, are not physical: they do not exist in general relativity.

Fortunately for us, nearby black holes or horizons are rare. Unfortunately, this means that neither the force limit nor the power limit are realized in any physical system at hand, neither at everyday length scales, nor in the microscopic world, nor in astrophysical systems. Even though the force and power limits have never been exceeded, a direct experimental confirmation of the limits will take some time.

The formulation of general relativity as a consequence of a maximum force is not Ref. 19 common; in fact, it seems that it was only discovered 80 years after the theory of general relativity had first been proposed.

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. 20 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 we 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

$$F = \frac{E}{L} . \tag{11}$$

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. 21 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, including those that are irregularly curved or time-dependent.[*]

Ref. 22 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 aspects of 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. Now, it is possible, by choosing a suitable coordinate transformation, to position a horizon at any desired space-time event. To achieve this, simply change to the frame of an observer accelerating away from that point at the correct distance, as explained in the volume on relativity. Vol. II, page 97 Therefore, because a horizon can be positioned anywhere at any time, 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 and power. Maximum force and general relativity are thus equivalent.

By the way, modern scholars often state that general relativity and gravity follow from the existence of a minimum measurable length. The connection was already stated by Ref. 23 Sakharov in 1969. This connection is correct, but unnecessarily restrictive. The maximum force, which is implicit in the minimal length, is sufficient to imply gravity. Quantum theory – or $\hbar$ – is (obviously) not necessary to deduce gravity.

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 Challenge 10 e $\sqrt{4GMa} \ll c^2$, 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 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 Challenge 11 e 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 gravity changes with the inverse square of distance. In other words, nature's limit on force implies universal gravitation. Other deductions of universal gravity from limit quantities are given Page 279 later.

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:

$$l \geqslant \frac{4Gm}{c^2} \; . \tag{20}$$

The size limit is only realized for *black holes*, those well-known systems which swallow everything that is thrown into them. The size limit is fully equivalent to the force limit. (Also the hoop conjecture is understood to be true.) All *composite* systems in nature comply with the lower size limit. Whether elementary particles fulfil or even match this Page 298 limit remains open at this point. More about this issue below.

General relativity also implies an 'indeterminacy relation' for the measurement errors Ref. 24 of size l and energy E of systems:

$$\frac{\Delta E}{\Delta l} \leqslant \frac{c^4}{4G} \; . \tag{21}$$

Experimental data are available only for composite systems; all known systems Ref. 25 comply with it. For example, the latest measurements for the Sun give $GM_{\odot}/c^3 = 4.925\,490\,947(1)\,\mu\mathrm{s}$; the error in E is thus much smaller than the (scaled) error in its radius, which is known with much smaller precision. The 'indeterminacy relation' (21) 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.

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. The importance of the maximum force 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 maximum force describes the elasticity of space-time. The high value of the maximum force tells us that it is difficult to bend space.

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 limit statements on motion:

$$\begin{aligned} &\text{Quantum theory follows from the action limit:} && W \geqslant \hbar \\ &\text{Thermodynamics follows from the entropy limit:} && S \geqslant k \\ &\text{Special relativity follows from the speed limit:} && v \leqslant c \\ &\text{General relativity follows from the force limit:} && F \leqslant \frac{c^4}{4G} . \end{aligned} \qquad (22)$$

These (corrected) *Planck limits* are valid for all physical systems, whether composite or elementary, and for all observers. Note that the limit quantities of quantum theory, thermodynamics, special 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 12 e

We note that the different dimensions of the four fundamental limits (22) in nature mean 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.

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

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

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

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

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

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

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

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

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

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

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

Of course, speed, action, angular momentum, entropy, power and force are also limited, as already stated. 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* or *corrected Planck limits* and assume that the numerical factor 4 has been properly included. In other words:

▷ Every natural unit or (corrected) Planck unit is the limit value of the corresponding physical observable.

Page 58 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. 26 The existence of a smallest measurable distance and time interval of the order of the Planck values is discussed in all approaches to quantum gravity. The maximum curvature Ref. 27 has been studied in quantum gravity; 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.

In the following, we often call the collection of Planck limits the *Planck scales*. We will discover shortly that at Planck scales, nature differs in many ways from what we are used to at everyday scales.

> 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 bindenden Kräften, gesucht werden.*
>
> Bernhard Riemann, 1854, *Über die Hypothesen, welche der Geometrie zu Grunde liegen.*

PHYSICS, MATHEMATICS AND SIMPLICITY

The four limits of nature of equation (22) – on action, entropy, speed and force – are astonishing. Above all, the four limits are *simple*. For many decades, a silent assumption has guided many physicists: physics requires *difficult* mathematics, and unification requires even more difficult mathematics.

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 difficult mathematics. This prejudice is a consequence of over a century of flawed teaching of physics. Flawed teaching is thus one of the reasons that the search for a final theory was not successful for so long.

The summary of physics with limit statements shows that nature and physics are *simple*. In fact, 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 with 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. 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 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.'

Bernhard Riemann is one of the most important mathematicians. 45 years after this statement, Max Planck confirmed that natural units are due to gravitation, and thus to 'binding forces'.

** Interestingly, he also regularly wrote the opposite, as shown on page 86.

*** Groucho Marx (b. 1890 New York City, d. 1977 Los Angeles), well-known comedian.

We have seen that 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 theories are kept separate. In other terms, 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. Formulating 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 58

Mass and energy limits

Mass plays a special role in all these arguments. The four limits (22) do 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 its size l is smaller than any measurable dimension. In particular, it must be smaller than the reduced Compton wavelength:

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

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

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

These elementary-particle limits are the (corrected) *Planck mass*, *Planck energy* and *Planck momentum*. They 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. 28

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 which move faster than light, which show actions below the quantum of action, and which experience forces larger than the force limit.

Vol. II, page 71 We know from special relativity that the virtual particles exchanged in collisions move faster than light. We know from quantum theory that the exchange of a virtual particle Vol. IV, page 191 implies actions below the minimum action. Virtual particles also imply an instantaneous change of momentum; they thus exceed the force limit.

In short, *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.

* *

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 research literature of the twentieth century.

* *

A further way to deduce the minimum length using the limit statements which structure this adventure 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 Vol. II, page 106 distance d can be expressed as

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

we see 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.*

---

* Sections called 'Curiosities' can be skipped at first reading.

* *

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

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

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 Planck-size horizons are expected to realize the temperature limit, but nobody has managed to explore them experimentally, so far.

* *

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 usually negligible. This is the reason why gravity must be included in a unified description of nature.

* *

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. Challenge 13 e

* *

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

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

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.

* *

It is well known that gravity bends space. Therefore, if they are to be fully convincing, our calculation for two attracting black holes 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, Ref. 29

the force would be

$$F = \frac{GMm}{d^2\sqrt{1-\frac{2GM}{dc^2}}} \,. \qquad (37)$$

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 $d = 2G(m + M)/c^2 \to 2GM/c^2$ requires $m \to 0$, 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 physical 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 14 e

* *

One possible system that actually achieves the Planck power limit is the final stage of black hole evaporation. But even in this case, the power limit is not exceeded. Challenge 15 e

* *

The maximum force limit states that the stress-energy tensor, when integrated over any physical surface, does not exceed the limit value. Ref. 19 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 realistic observer, in particular, by an observer 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.

* *

The power limit is of interest if applied to the universe as a whole. Indeed, it can be used to partly 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 35 Challenge 16 s 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 from ripping apart. Does this analogy make sense?

* *

Ref. 23 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 at small distances will induce counter-terms that include to lowest order the cosmological constant and then the Einstein–Hilbert action of general relativity.

* *

We said above that a surface is physical if an observer can be attached to each of its points. The existence of a smallest length – and a corresponding shortest time interval – implies

▷ No surface is *physical* if any part of it requires a localization in space-time to scales below the minimum length.

For example, a physical surface must not cross any horizon. Only by insisting on physical surfaces 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. 30

* *

The equation $E = c^2 m$ implies that energy and mass are equivalent. What do the equations $l = (4G/c^2)m = (4G/c^4)E$ for length and $W = \hbar\varphi$ for action imply? Challenge 17 e

* *

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

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

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

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

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

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

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

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

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

For the case of a single conduction channel, we get

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

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

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

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.

Ref. 31

Ref. 32

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 relation that to every limit value in nature there is a corresponding indeterminacy relation is valid also 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. 33 Both these relations may be found in the research 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 motion 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 lower 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 106

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 18 e

$$D_{\text{system}} \leqslant \frac{1}{c} LE . \tag{50}$$

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. 34

---

* 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 149

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. By the way, 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. Though nature always seeks to minimize the action W, systems, of size L, that *maximize* action are also interesting. You might check for yourself that the action limit

$$W \leqslant LE/c \qquad (51)$$

Challenge 19 e 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:

$$v_{\text{system}} \geqslant \hbar c^2 \frac{1}{LE} \ . \qquad (52)$$

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. Challenge 20 e

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 – just the global limit c. Challenge 21 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. Challenge 22 s

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

$$F_{\text{system}} \geqslant \frac{\hbar}{2c} \frac{1}{T^2} \ . \qquad (53)$$

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

$$c^3 \frac{M}{L} \geqslant P_{\text{system}} \geqslant 2\hbar G \frac{M}{L^3} \ . \qquad (54)$$

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 research literature are not yet consistent.)

THE STRANGE CHARM OF THE ENTROPY BOUND

Ref. 35 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

$$\frac{S}{S_{\text{c.Planck}}} \leqslant \frac{A}{A_{\text{c.Planck}}} \tag{55}$$

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.

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

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

This is called *Bekenstein's entropy bound*. It states that the entropy of any physical system is *finite* and limited by its mass M and size L. No exception has ever been found or constructed, despite many attempts. Again, the limit value itself is only realized for black holes. Ref. 30

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) we therefore have 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 Ref. 36 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

$$S \leqslant \frac{k}{\hbar}\eta V \; . \tag{59}$$

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 23 e

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 finite 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 24 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 25 s

* *

Challenge 26 e Can you find a general method to deduce all limits of observables?

* *

Bekenstein's entropy bound 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

Challenge 27 e

number $N_{\text{d.o.f.}}$ of degrees of freedom in the universe is roughly

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

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 far 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 following idea: 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

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

This is the smallest temperature that a system of mass M and size L can have. 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 the claim on the minimum system temperature is known. All systems from everyday life comply with it, as do all stars. Also the coldest known systems in the universe, namely Bose–Einstein condensates and other cold gases 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 28 s

Vol. II, page 61

By the way, there seems to be no consistent way to define an upper limit for a size-dependent temperature. Limits for other thermodynamic quantities can be found, but

Challenge 29 ny

we do not 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 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 might 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. 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.

Challenge 30 s

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. Does the size limit change when electric charge is taken into account? In fact, an entire research field is dedicated to deducing and testing the most general limits valid in nature.

* *

Many cosmological limits have not been discussed here nor anywhere else. The following could all be worth a publication: What is the limit for momentum? Energy? Pressure? Acceleration? Mass change? Lifetime?

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 stating

▷ There is a maximum distance value of the order of $1.4/\sqrt{\Lambda}$ in nature.

For all systems and all observers, sizes, distances and lengths are limited by the relation

$$l \lesssim \frac{1.4}{\sqrt{\Lambda}} = 1.3 \times 10^{26}\,\text{m} = 1.4 \times 10^{10}\,\text{al} . \tag{66}$$

This expression contains all of cosmology. The details of the numerical factor 1.4 are not of importance here and we will often neglect it in the following. This statement on length should be added to the four fundamental Planck limits as a fifth limit statement in nature.

Challenge 31 s 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.

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

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

It has never been reached or approached by any observation.

Many cosmological limits are related to black hole limits. 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 32 e

Minimum force

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.

In summary,

▷ Nature provides two limits for each observable: a Planck limit and a cosmological limit.

Every observable has a lower and an upper limit. You may want to summarize them into a table. This has important consequences that we will explore now.

Challenge 33 s

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.

Whenever 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 scales, 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 58

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. 37

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.

Page 81

MEASUREMENT PRECISION AND THE EXISTENCE OF SETS

In physics, it is generally assumed that nature is a *set* of components or parts. 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.

The existence of 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 58

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 possible. In fact, even at everyday energies, any distinction between two physical systems – for example, between a toothpick and a mountain – is possible only *approximately*. At Planck scales, a boundary can never be drawn.

Page 58

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.

Page 58

▷ Nature has no parts.

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 parts or entities.* In nature, elements of sets cannot be defined. Neither discrete nor continuous sets can be constructed:

Page 58

▷ Nature does not contain sets or elements.

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 – particles, space, time, observables, phase space, 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.

- ▷ No correct mathematical model of nature can be based on sets.

In other terms, nature has no parts: nature is one.

None of the approaches to unification pursued in the twentieth century has abandoned sets. This requirement about the final theory is thus powerful and useful. Indeed, the requirement to abandon sets will be an efficient guide in our search for the unification of relativity and quantum theory. The requirement will even solve Hilbert's sixth problem. Page 108

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 is equivalent to general relativity.
- ▷ The action limit is equivalent to quantum theory.
- ▷ The entropy limit is equivalent to thermodynamics.
- ▷ The distance limit is equivalent to cosmology.

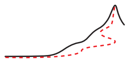
All these limits are observer-invariant. The invariance of the limits suggests interesting thought experiments, none of which leads to their violation.

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 of points is not correct, and nature can be described by sets and parts only approximately. At Planck scales, nature does not contain sets or elements.

Nature's limits imply that Planck units are the key to the final theory. 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 allow an equally simple description. Nature's limits thus suggest that the mathematics of the final, unified theory might be *simple*.

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 that Page 19 is simple and *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 accurate 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. This millennium description of motion is *useful* because its consequences are confirmed by all experiments, to the full measurement precision. 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. Nevertheless, if a physicist has some other concept of beauty in physics, avoid him, because in that case he is really talking nonsense.

The simplicity of twentieth-century physics is well-known: 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*. In particular, every example of motion due to general relativity or to the standard model of particle physics minimizes action: both theories can be described concisely with the help of a Lagrangian.

Vol. I, page 245

On the other hand, some important aspects 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.

The remaining part of our adventure 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, useful and simple, we do carry a burden: the two theories and their concepts *contradict* each other.

[**] ‘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 complete vacuum.

In everyday life, vacuum has an energy density that cannot be distinguished from zero. However, general relativity proposes a way to check this with high precision: we measure the average curvature of the universe. Nowadays, cosmological measurements performed with dedicated satellites reveal an average energy density E/V of the intergalactic 'vacuum' with the value of

Ref. 38

Vol. II, page 242

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

In short, cosmological data show that the energy density of intergalactic space is not exactly zero; nevertheless, the measured value is extremely small and can be neglected in all laboratory experiments.

Ref. 39

On the other hand, quantum field theory tells a different story on vacuum energy density. 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. Following quantum field theory, the most precise theory known, their energy density is given, within one order of magnitude, by

Vol. V, page 121

Vol. V, page 127

$$\frac{E}{V} \approx \frac{4\pi h}{c^3}\int_0^{\nu_{\max}} \nu^3 \mathrm{d}\nu = \frac{\pi h}{c^3}\nu_{\max}^4\ . \qquad (70)$$

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 $0.6 \cdot 10^{19}$ GeV= 1.0 GJ. That would give a vacuum energy density of

Page 39

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

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-time. 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

Ref. 41

---

* It is worthwhile to stress that the 'slight' mistake lies in the domain of quantum field theory. There is no mistake and no mystery, despite the many claims to the contrary found in newspapers and in bad research articles, in general relativity. This well-known point is made especially clear by Bianchi and Rovelli.

Ref. 40

in quantum field theory. In short, quantum theory only works because it assumes that gravity does not exist. Indeed, the gravitational constant G does not appear in quantum field theory.

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.

The concept of *measurement* also differs. 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 relation yields limits that follow from the mass M of the measurement apparatus.

Vol. V, page 289

Vol. V, page 43

The contradictions also concern the concept of *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? It means nothing.

Page 64

Similarly, general relativity implies that space and time cannot be distinguished, whereas quantum theory implies that matter does make a distinction between them. A related difference is the following. 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 283

The contradiction between the two theories is shown most clearly 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*, but fermions cannot be incorporated into general relativity.[**]

Ref. 42, Ref. 43

Ref. 44, Ref. 45

Ref. 46

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 or to a completely correct description. To eliminate the contradictions, we need to understand their origin.

THE ORIGIN OF THE CONTRADICTIONS

Ref. 47

All contradictions between general relativity and quantum mechanics have the same origin. In 20th-century physics, motion is described in terms of objects, made up of

---

[*] John Archibald Wheeler (b. 1911, Jacksonville, d. 2008, Hightstown), was a physicist and influential teacher who worked on general relativity.

[**] As we will see below, the strand model provides a way to incorporate fermions into an extremely accurate approximation of general relativity, without requiring any topology change. This effectively invalidates Wheeler's argument.

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 that has a position and that 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. 48 An *event* is a point in space and time. 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 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: in the millennium description of nature, *the two concepts of particle and of space-time are each defined with the help of the other*. This circular definition is the origin of the contradictions between quantum mechanics and general relativity. In order to eliminate the contradictions and to formulate a complete theory, we must eliminate this circular definition.

THE DOMAIN OF CONTRADICTIONS: PLANCK SCALES

Despite their contradictions and the underlying circular definition, both general relativity and quantum theory are successful theories for the description of nature: they agree with all data. How can this be?

Each theory of modern physics provides a criterion for determining when it is necessary and when classical Galilean physics is no longer applicable. These criteria are the basis for many arguments in the following chapters.

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 unne- Ref. 42, Ref. 49

---

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

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 (red.)	Ratio d/λ_C
galaxy	$\approx 1\,\mathrm{Zm}$	$\approx 5 \cdot 10^{40}\,\mathrm{kg}$	$\approx 70\,\mathrm{Tm}$	$\approx 10^{7}$	‘$\approx 10^{-83}\,\mathrm{m}$’	$\approx 10^{104}$
neutron star	$10\,\mathrm{km}$	$2.8 \cdot 10^{30}\,\mathrm{kg}$	$4.2\,\mathrm{km}$	2.4	‘$1.3 \cdot 10^{-73}\,\mathrm{m}$’	$8.0 \cdot 10^{76}$
Sun	$1.4\,\mathrm{Gm}$	$2.0 \cdot 10^{30}\,\mathrm{kg}$	$3.0\,\mathrm{km}$	$4.8 \cdot 10^{5}$	‘$1.0 \cdot 10^{-73}\,\mathrm{m}$’	$8.0 \cdot 10^{81}$
Earth	$13\,\mathrm{Mm}$	$6.0 \cdot 10^{24}\,\mathrm{kg}$	$8.9\,\mathrm{mm}$	$1.4 \cdot 10^{9}$	‘$5.8 \cdot 10^{-68}\,\mathrm{m}$’	$2.2 \cdot 10^{74}$
human	$1.8\,\mathrm{m}$	$75\,\mathrm{kg}$	$0.11\,\mathrm{ym}$	$1.6 \cdot 10^{25}$	‘$4.7 \cdot 10^{-45}\,\mathrm{m}$’	$3.8 \cdot 10^{44}$
molecule	$10\,\mathrm{nm}$	$0.57\,\mathrm{zg}$	‘$8.5 \cdot 10^{-52}\,\mathrm{m}$’	$1.2 \cdot 10^{43}$	$6.2 \cdot 10^{-19}\,\mathrm{m}$	$1.6 \cdot 10^{10}$
atom (^{12}C)	$0.6\,\mathrm{nm}$	$20\,\mathrm{yg}$	‘$3.0 \cdot 10^{-53}\,\mathrm{m}$’	$2.0 \cdot 10^{43}$	$1.8 \cdot 10^{-17}\,\mathrm{m}$	$3.2 \cdot 10^{7}$
proton p	$2\,\mathrm{fm}$	$1.7\,\mathrm{yg}$	‘$2.5 \cdot 10^{-54}\,\mathrm{m}$’	$8.0 \cdot 10^{38}$	$2.0 \cdot 10^{-16}\,\mathrm{m}$	9.6
pion π	$2\,\mathrm{fm}$	$0.24\,\mathrm{yg}$	‘$3.6 \cdot 10^{-55}\,\mathrm{m}$’	$5.6 \cdot 10^{39}$	$1.5 \cdot 10^{-15}\,\mathrm{m}$	1.4
up-quark u	$< 0.1\,\mathrm{fm}$	$5 \cdot 10^{-30}\,\mathrm{kg}$	‘$7 \cdot 10^{-57}\,\mathrm{m}$’	$< 1 \cdot 10^{40}$	$7 \cdot 10^{-14}\,\mathrm{m}$	< 0.001
electron e	$< 4\,\mathrm{am}$	$9.1 \cdot 10^{-31}\,\mathrm{kg}$	‘$1.4 \cdot 10^{-57}\,\mathrm{m}$’	$< 3 \cdot 10^{39}$	$3.9 \cdot 10^{-13}\,\mathrm{m}$	$< 1 \cdot 10^{-5}$
neutrino ν_e	$< 4\,\mathrm{am}$	$< 3 \cdot 10^{-36}\,\mathrm{kg}$	‘$< 5 \cdot 10^{-63}\,\mathrm{m}$’	n.a.	$> 1 \cdot 10^{-7}\,\mathrm{m}$	$< 3 \cdot 10^{-11}$

cessary in everyday life. We recall that objects whose size is given by their Schwarzschild radius are black holes; smaller objects cannot exist.

Challenge 34 e

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

$$\bar{\lambda}_C = \frac{\hbar}{m c} \,. \tag{73}$$

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 pair 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.

Combining concepts of quantum field theory and general relativity is 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 – within a factor of order 1 – of the order of

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

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.

Challenge 35 e

In the millennium description of nature, all the contradictions and also the circular definition just mentioned are effective only at Planck scales. You can check this yourself. This is the reason that general relativity and quantum theory work so well in practice.

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 Planck scales.

In summary, general relativity and quantum theory do contradict each other. However, the domains where these contradictions play a role, the Planck scales, are not accessible by experiment. As a consequence, the contradictions and our lack of knowledge of how nature behaves at the Planck scales have only one effect: we do not see the solutions to the millennium issues.

We note that some researchers argue that the Planck scales specify only one of several domains of nature where quantum mechanics and general relativity apply simultaneously. They mention horizons and the big bang as separate domains. However, it is more appropriate to argue that horizons and the big bang are situations where Planck scales are essential.

RESOLVING THE CONTRADICTIONS

The contradictions between general relativity and quantum theory have little practical consequences. Therefore, for a long time, the contradictions were accommodated by keeping the two theories separate. It is often said that quantum mechanics is valid at small scales and general relativity is valid at large scales. This attitude is acceptable as long as we remain far from the Planck length. However, this accommodating attitude also prevents us from resolving the circular definition, the contradictions and therefore, the millennium issues.

The situation resembles the well-known drawing, Figure 2, by Maurits Escher (b. 1898 Leeuwarden, d. 1972 Hilversum) 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. In physics, the third concept is the common origin of vacuum and particles.

We thus conclude that the contradictions in physics and the circular definition are solved by *common constituents* for vacuum and matter. In order to find out what these common constituents are and what they are not, we must explore the behaviour of nature at the Planck scales.

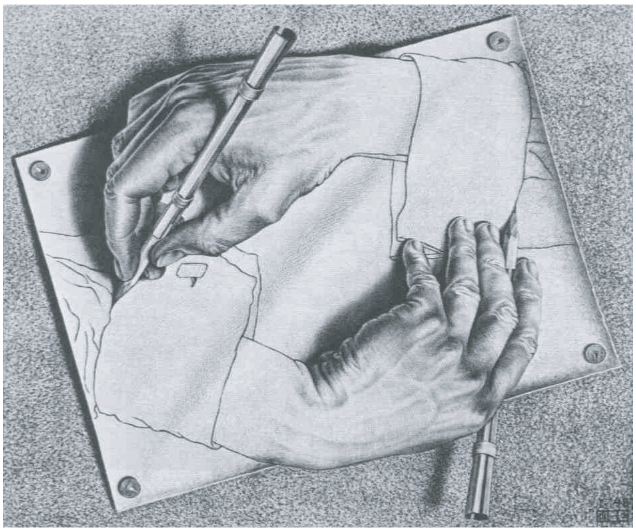


FIGURE 2 'Tekenen' by Maurits Escher, 1948 – a metaphor for the way in which 'particles' and 'space-time' are defined: each with the help of the other (© M.C. Escher Heirs).

The origin of points

General relativity is built on the *assumption* that space is a continuum of points. Already at school we learn that lines, surfaces and areas are made of points. We take this as granted, because we imagine that finer and finer measurements are always possible. And all experiments so far agree with the assumption. Fact is: in this reasoning, we first idealized measurement rulers – which are made of matter – and then 'deduced' that points in space exist.

Quantum theory is built on the *assumption* that elementary particles are point-like. We take this as granted, because we imagine that collisions at higher and higher energy are possible that allow elementary particles to get as close as possible. And all experiments so far agree with the assumption. Fact is: in this reasoning, we first imagined infinite energy and momentum values – which is a statement on time and space properties – and then 'deduce' that point particles exist.

▷ The use of points in space and of separate, point-like particles are the reasons for the mistaken vacuum energy calculation (71) that is wrong by 120 orders of magnitude.

Page 57

In short, only the circular definition of space and matter allows us to define points and point particles. This puts us in a strange situation. On the one hand, experiment tells us that describing nature with space points and with point particles works. On the other hand, reason tells us that this is a fallacy and cannot be correct at Planck scales. We need a solution.

SUMMARY ON THE CLASH BETWEEN THE TWO THEORIES

General relativity and quantum theory contradict each other. In practice however, this happens only at Planck scales. The reason for the contradiction is our insistence on a circular definition of space and particles. Indeed, we need this circularity: Only such a circular definition allows us to define *points* and *point particles* at all.

In order to solve the contradictions between general relativity and quantum theory and in order to understand nature at Planck scales, we must introduce *common* constituents for space and particles. But common constituents have an important consequence: common constituents force us to stop using points to describe nature. We now explore this connection.

Chapter 4

DOES MATTER DIFFER FROM VACUUM?

Vol. II, page 23

The appearance of the quantum of action in the description of motion leads to limitations for all measurements: Heisenberg's indeterminacy relations. These relations, when combined with the effects of gravitation, imply an almost unbelievable series of consequences for the behaviour of nature at Planck scales. The most important ones are the necessity to abandon points, instants and events, and the equivalence of vacuum and matter. Here we show how these surprising and important conclusions follow from simple arguments based on the indeterminacy relations, the Compton wavelength and the Schwarzschild radius.

Farewell to instants of time

> Time is composed of time atoms ... which in fact are indivisible.
>
> Maimonides[**]

Ref. 50, Ref. 51

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. 52, Ref. 53
Ref. 54

Every clock is a device with some moving parts. The moving parts can be mechanical wheels, or particles of matter in motion, or changing electrodynamic fields (i.e., photons), or decaying radioactive particles. For each moving component of a clock the indeterminacy relation 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

---

[**] Moses Maimonides (b. 1135 Cordoba, d. 1204 Egypt) was a physician, philosopher and influential theologian. However, there is no evidence for 'time atoms' in nature, as explained below.

Page 74

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. Now, ΔE must be smaller than the total energy $E = c^2 m$ of the component itself: $\Delta E < c^2 m$.* 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 a black hole of 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: $l < c\delta t$.** If we combine these three conditions, we get

$$\delta t \geqslant \frac{\hbar G}{c^5 \delta t} \qquad (76)$$

or

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

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 162) 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. 49). From the mass–energy relation, we see that an energy spread

---

* 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 36 s)

*** 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. 55, Ref. 56). In our case, both quantum mechanics and gravity are included, and therefore a different, lower, and more fundamental limit is found. Also the discovery of black hole radiation does not change the argument (Ref. 57, Ref. 58): black hole radiation notwithstanding, measurement devices cannot exist inside black holes.

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 $\Delta E \geqslant \hbar/\Delta t$. 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, it is impossible to measure time intervals shorter than the Planck time. Thus

▷ 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 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.* Ref. 60

FAREWELL TO POINTS IN SPACE

> Our greatest pretenses are built up not to hide the evil and the ugly in us, but our emptiness. 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 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. Ref. 24

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

---

* Eric Hoffer (b. 1902 New York City, d. 1983 San Francisco), philosopher.

must be larger than the position indeterminacy of the two objects. From the indeterminacy relation we know that neither object's position can be determined with a precision Δl better than that given by $\Delta l \, \Delta p = \hbar$, 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

$$\delta l \geqslant \Delta l \geqslant \frac{\hbar}{mc} \; . \qquad (79)$$

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

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

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 $\Delta E = c(c^2 m^2 + (\Delta p)^2)^{1/2} \geqslant c\hbar/\Delta x$. 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. 42, Ref. 49
Ref. 24
Ref. 61
Ref. 62, Ref. 63
Ref. 26

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 $V_{\mathrm{Pl}} = l_{\mathrm{Pl}}^3$.

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 or theoretical basis*. In other terms,

▷ In nature there is a minimum length interval.

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 RELATION

Ref. 24 The limit values for length and time measurements are often expressed by the so-called *generalized indeterminacy relation*

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

or

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

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 37 e

The generalized indeterminacy relation is derived in exactly the same way in which Heisenberg derived the original indeterminacy relation $\Delta p \Delta x \geqslant \hbar/2$, namely by studying the scattering 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 all known approaches do so. Ref. 24, Ref. 64, Ref. 65, Ref. 66, Ref. 67, Ref. 68

FAREWELL TO SPACE-TIME CONTINUITY

> “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.

The classical description of nature is based on continuity: it involves and allows differences of time and space that are as small as can be imagined. Between any two points

---

[*] '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.'

in time or space, the existence of infinitely many other points is assumed. Measurement results of arbitrary small values are deemed possible. The same is valid for action values.

However, 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. Therefore, *the continuum description of space-time has to be abandoned* in favour of a more appropriate description.

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 Ref. 69

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

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. 70

> 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.*

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.

▷ 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

---

* '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 for our present means of description failing 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!

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.'

Page 86 Compare this letter to what Einstein wrote almost twenty and almost forty years later.

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. 24, Ref. 71

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 coordinates 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

---

* 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.

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

Ref. 72

The reasons for the problems with space-time become most evident when we remember 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.)

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 56
Ref. 73

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.

---

* 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.

▷ 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.

Farewell to observables, symmetries 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.* Quantum mechanics also relies on the possibility to add wave functions; this is sometimes called the *superposition principle*. Without fields and superpositions, all of quantum mechanics comes crumbling down.

The lack of real numbers has severe consequences. 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.* 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 superpositions cannot be backed up by experiment, even the famous Wheeler–DeWitt equation, sometimes 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

Vol. V, page 358

---

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

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, because of the existence of a smallest length. As a result, the *renormalization group breaks down at Planck scales.*

In summary, due to the impossibility of accurate measurements, all symmetries break down at Planck scales. (For example, supersymmetry cannot be valid at Planck scale.) 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 OR SPACE-TIME BE A LATTICE?

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

Ref. 74 Ref. 75 Ref. 76 Discrete models of space-time have been studied since the 1940s. Recently, the idea that space or space-time could be described as a lattice – like a crystal – has been explored most notably by David Finkelstein and by Gerard 't Hooft. The idea of space as a lattice is based on the idea that, if there is a minimum distance, then all distances are multiples of this minimum. Ref. 77

Ref. 78 In order to get an isotropic and homogeneous situation for large, everyday scales, the structure of space 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, *neither space nor space-time can be a lattice.* A minimum distance does exist in nature; however, we cannot hope that all other distances are simple multiples of it.

▷ Space is not discrete. Neither is space-time.

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 *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. 27 Several research approaches 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. The naive statement is: a particle is elementary if it behaves like a point particle. At present, only the leptons (electron, muon, tau and the neutrinos), the quarks, the radiation quanta of the electromagnetic, weak and strong nuclear interactions (the photon, the W and Z bosons, and the gluons) and the Higgs boson have been found to be elementary. Protons, atoms, molecules, cheese, people, galaxies and so on are all composite, as shown in Table 2. Vol. V, page 260 Page 60

Although the naive definition of 'elementary particle' as point particle is all we need in the following argument, the definition is not precise. It seems to leave open the possibility that future experiments could show that electrons or quarks are not elementary. This is not so! In fact, the precise definition is the following:

Vol. IV, page 105
▷ Any particle is elementary if it is *smaller than its own Compton wavelength*.

If such a small particle 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, for the size d of an object, the relation

$$d > \sqrt{\frac{\hbar G}{c^3}} = l_{\mathrm{Pl}} \; . \tag{84}$$

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

▷ There is 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.*

The minimum size for particles can be tested. A property connected with the size is the 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. 80

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

where e is the charge of the electron. This predicted value is ten thousand times smaller than the Planck length l_{Pl}. Since the Planck length is the smallest possible length, we

---

* 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.

Ref. 81 seem to have a contradiction here. However, a more careful and recent prediction from the standard model only states

$$\frac{|d_e|}{e} < 3 \cdot 10^{-23}\,\mathrm{m}\,, \tag{86}$$

Ref. 82 which is not in contradiction with the minimal length. The experimental limit in 2013 is

$$\frac{|d_e|}{e} < 8.7 \cdot 10^{-31}\,\mathrm{m}\,. \tag{87}$$

In the coming years, the experimental limit value will approach the Planck length. In summary, no point particle is known. In fact, not even a particle smaller than the Planck length is known.

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. 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.

And we can continue. Due to measurement limitations, also *spatial parity* cannot be defined or measured at Planck scales. Challenge 38 e

We have thus shown that at Planck scales, particles do not interact locally, are not

FIGURE 3 Andrei Sakharov (1921–1989).

point-like, cannot be distinguished from antiparticles, cannot be distinguished from virtual particles, have no definite spin and have no definite spatial parity. We deduce 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: $d > l_{\text{Pl}}$. Combining these two requirements and eliminating the size d, we get a constraint on the mass m of any elementary particle, namely

$$m < \frac{\hbar}{c\, l_{\text{Pl}}} = \sqrt{\frac{\hbar c}{G}} = m_{\text{Pl}} = 2.2 \cdot 10^{-8}\,\text{kg} = 1.2 \cdot 10^{19}\,\text{GeV}/c^2 \,. \qquad (88)$$

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 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.) Ref. 28

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 $m < m_{\text{Pl}}$. 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. 83

---

* Andrei Dmitrievich Sakharov, Soviet nuclear physicist (b. 1921 Moscow, d. 1989 Moscow). 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.

Farewell to massive particles – and to massless vacuum

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

$$\rho_{\rm Pl} = \frac{c^5}{G^2\hbar} = 5.2 \cdot 10^{96}\,\mathrm{kg/m^3} \tag{89}$$

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 $M < c^2R/G$; 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 39 e

$$\rho < \rho_{\rm Pl} \ . \tag{90}$$

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

$$\Delta M \geqslant \frac{\hbar}{cR} \ . \tag{91}$$

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 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 (91). Now that we are sure of the result, let us continue.

From equation (91) 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: $\Delta M \geqslant M_{\rm Pl}$. 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

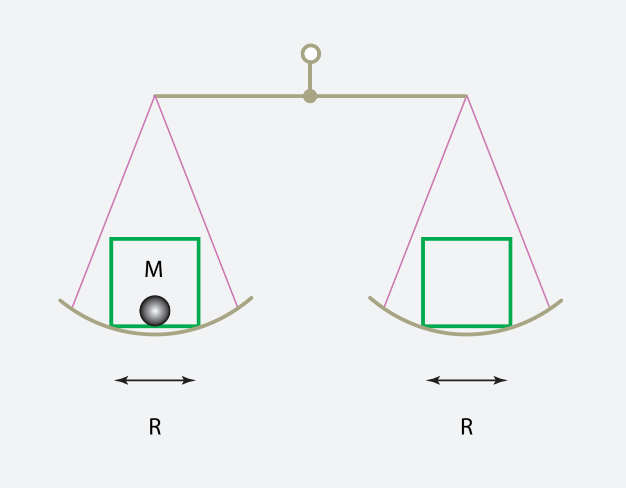


FIGURE 4 A thought experiment showing that matter and vacuum cannot be distinguished when the size of the enclosing box is of the order of a Planck length.

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:

$$\underset{\text{(full box)}}{\frac{c^2R}{G}} \geqslant m \geqslant \underset{\text{(empty box)}}{\frac{\hbar}{cR}} . \tag{92}$$

We see directly that for sizes R of the order of the Planck scales, 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

$$\frac{\Delta M}{M} = \frac{\Delta\delta\lambda}{\delta\lambda} . \tag{93}$$

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 $l_{\rm Pl}$. In other words, for a Planck volume the wavelength error – and thus the mass error – is always as large as the Planck mass itself: $\Delta M \geqslant M_{\rm Pl}$. 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', of the same constituents. 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 the Fulling–Davies–Unruh effect. 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. 84

Vol. V, page 127

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 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.* In particular, *the concept of particle is not applicable at Planck scale.*

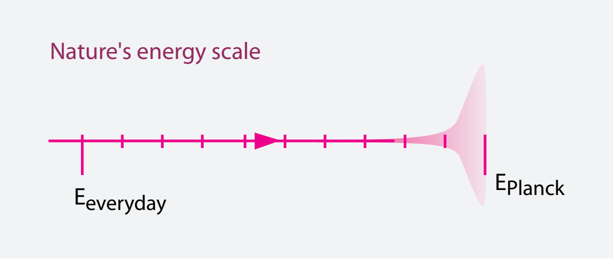


FIGURE 5 Planck effects make the energy axis an approximation.

CURIOSITIES AND FUN CHALLENGES ON PLANCK SCALES

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

Lenin’s statement is wrong. And this is not so much because the world contains moving matter, moving radiation, moving vacuum and moving horizons, which is not exactly what Lenin claimed. Above all, his statement is wrong because at Planck scales, there is no matter, no radiation, no horizon, no space and no time. These concepts only appear at low energy. The rest of our adventure clarifies how.

* *

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 scales, vacuum, radiation and matter cannot be distinguished. Two consequences follow: first, only at Planck scales would a description be free of any bias in favour of matter. Secondly, on the other hand, observers do not exist at all at Planck energy. Physics is thus only possible below 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.

* *

Challenge 40 s

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 41 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?

* *

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.*

* *

Challenge 42 s Do the large errors in mass measurements imply that mass can be negative at Planck 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 Challenge 43 r follow? What kind of experiment could measure them? If you have a suggestion, Ref. 85 publish it! Several attempts are being explored. We will settle the issue later on, with some Page 277 interesting insights.

* *

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., Ref. 86 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 Ref. 67 single atoms inside the crystal.

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 Challenge 44 s incorrect, at least for space and time?

* *

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 Challenge 45 s energy is the highest energy in nature. What does this apparent contradiction imply?

* *

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

* *

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. II, page 273

* *

Because mass and energy density are limited, any object of finite volume has only a finite number of degrees of freedom. The calculation of the entropy of black holes has confirmed 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. V, page 144 Ref. 87

* *

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

* *

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 interactions are exchanges of quantum particles. This apparent contradiction can be used to show that something is counted twice in the usual characterization of nature. Noting that matter and space-time are both made of the same constituents resolves the issue. Vol. III, page 317

* *

Is there a smallest possible momentum? And a smallest momentum error? Challenge 46 d

* *

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 47 s Page 408 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.

* *

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 115

* *

The impossibility of distinguishing 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 285

* *

Challenge 48 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.

* *

If the dimensionality of space is undefined at Planck scales, what does this mean for superstrings? Challenge 49 e

* *

Since vacuum, particles and fields are indistinguishable at Planck scales, we also lose the distinction between states and permanent, intrinsic properties of physical systems at those scales. 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. 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. Vol. I, page 26

COMMON CONSTITUENTS

> “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-Kontinuums 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*.

> “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.

In this rapid journey, we have destroyed all the experimental pillars of quantum theory: the superposition of wave functions, 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? Page 68

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.

Secondly, after tearing down the castles, the invariant Planck limits c, $\hbar$ and $c^4/4G$ still remain as a foundation.

Thirdly, after tearing down the castles, one important result appears. 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

---

* ‘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.’

Page 69 See also what Einstein thought twenty years before. The new point is that he believes that an algebraic description is necessary. He repeats the point in the next quote.

and particles must be made of *common constituents*. In other words, we are not in airless space, and we uncovered the foundation that remains after we tore down the castles. Before we go on exploring these common constituents, we check what we have deduced so far against experiment.

Experimental predictions

A race is going on both in experimental and in theoretical physics: to be the first to suggest and to be the first to perform an experiment that detects a quantum gravity effect – apart possibly from (a part of) the Sokolov–Ternov effect. Here are some proposals.

Challenge 50 r

Vol. V, page 145

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. However, no blurring is observed, and the first tests show that light from extremely distant galaxies still interferes. The precise prediction of the phase washing effect is still being worked out; whatever the exact outcome, the effect is too small to be measured. Ref. 88

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 light signals when they travel across the universe. Ref. 92, Ref. 91 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

$$E_{\text{char}} = \frac{(E_1 - E_2)\, d}{c\, \Delta t} \;. \qquad (94)$$

This energy value is between $1.4 \cdot 10^{19}$ GeV and over 10^{22} GeV for the best measurement to date. Ref. 89, Ref. 90 This is between just above the Planck energy and over one thousand times the Planck energy. Ref. 90 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 bodies. Ref. 92, Ref. 93 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 to be Ref. 94

$$\frac{\delta l}{l} \geqslant \left(\frac{l_{\text{Pl}}}{l}\right)^{2/3} \;. \qquad (95)$$

This expression is deduced simply by combining the measurement limit of a ruler, from quantum theory, with the requirement that the ruler not be a black hole. Page 66 The sensitivity of the detectors to noise might reach the required level in the twenty-first century. The noise induced by quantum gravity effects has also been predicted to lead to detectable quantum decoherence and vacuum fluctuations. Ref. 95 So far, no such effect has been found.

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 Ref. 92 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.

Another possibility is that quantum gravity effects may change the threshold energy at Ref. 96 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 at Planck scales. However, no such effect has been found yet.

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 Ref. 95 transitions. However, no such effect has been found yet.

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.

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 prediction and 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 relativ- Ref. 97 ity did. On the other hand, most unified models of physics predict the *absence* of any measurable quantum gravity effect.

SUMMARY ON PARTICLES AND VACUUM

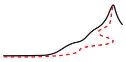
Combining quantum theory and general relativity leads us to several important results on the description of nature:

- *Vacuum and particles mix* at Planck scales, because 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. Equivalently, empty Ref. 98 space and particles are made of *fluctuating common constituents*.
- We note that all arguments of this chapter equally imply that vacuum and particles mix *near* Planck scales. For example, matter, radiation and vacuum cannot be distinguished *near* Planck scales.
- The constituents of vacuum and particles *cannot be points*. There is no conceivable way to prove that points exist, because the smallest measurable distance in nature is the Planck length.
- Particles, vacuum and continuous space *do not exist* at Planck scales. They disappear in a yet unclear Planck scale mixture.
- The three independent Planck limits c, $\hbar$ and $c^4/4G$ *remain valid* also in domains where quantum theory and general relativity are combined.

All these results must be part of the final theory that we are looking for. Generally speak-

Page 54

ing, we found the same conclusions that we found already in the chapter on limit statements. We thus continue along the same path that we took back then: 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. In fact, 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

$$d \approx r_S = 2Gm/c^2 \ . \tag{96}$$

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. Challenge 51 s

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 (reduced) Compton wavelength $\lambda\!\!\!^{-}_C$, i.e., whenever

$$d \approx \lambda\!\!\!^{-}_C = \frac{\hbar}{mc} \ . \tag{97}$$

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 (b. 1886 Starzeddel, d. 1965 Chicago), theologian, socialist and philosopher.

[***] 'Here are lions.' This was 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

$$13\,800 \text{ million years, or } 435\,\text{Ps},$$

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.

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. Now, 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.

Challenge 52 s

In summary, it is not possible to measure time intervals greater than the maximum time, 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 age 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**

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?

Vol. II, page 307

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. How-

---

* 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.

** Oscar Wilde, (b. 1854 Dublin, d. 1900 Paris), poet and playwright, equally famous for his wit.

ever, 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, no clock can run through its earliest history.

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 possibly 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. Despite this correspondence, for the universe as a *whole*, an age *cannot* be defined, because there is no clock outside it!

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.*

The lack of a beginning is consistent with a minimum length or a minimum action. Indeed, 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

$$K < \frac{c^3}{G\hbar} = 0.39 \cdot 10^{70}\,\mathrm{m}^{-2} \qquad (98)$$

as a limit for the surface curvature K in nature. In other words, the universe has never been as small as 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, Challenge 53 s

▷ The big bang cannot have been an event.

There never was an initial singularity or a beginning of the universe.

In short, 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? To clarify the issue, we 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. 55, Ref. 56

$$m > \frac{\hbar}{c^2} \frac{T}{(\Delta t)^2} . \tag{99}$$

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. 94

$$\Delta t > \frac{Gm}{c^3} . \tag{100}$$

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

$$T < \frac{(\Delta t)^3}{t_{\text{Pl}}^2} , \tag{101}$$

where $t_{\text{Pl}} = \sqrt{\hbar G/c^5} = 5.4 \cdot 10^{-44}$ 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 54 e

As a consequence of the inequality (101), 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 (101), 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.

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.

DOES TIME EXIST?

> “Time is waste of money.”
>
> Oscar Wilde

---

* 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 properties 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. 99

Ever since people began to study physics, the concept of 'time' has designated what is measured by a clock. But the inequality (101) 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 observable or any everyday concept. Vol. I, page 39

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. Challenge 55 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. Page 58 Page 68 Challenge 56 e

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?

It is now straightforward to apply our discussion about the measurement of time to the age of the universe. The inequality (101) implies that the highest precision possible for a clock is about $10^{-23}\,\mathrm{s}$, 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 (101) can be written as Challenge 57 e

$$\frac{\Delta t}{T} > \left(\frac{t_{\mathrm{Pl}}}{T}\right)^{2/3} . \qquad (102)$$

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.

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

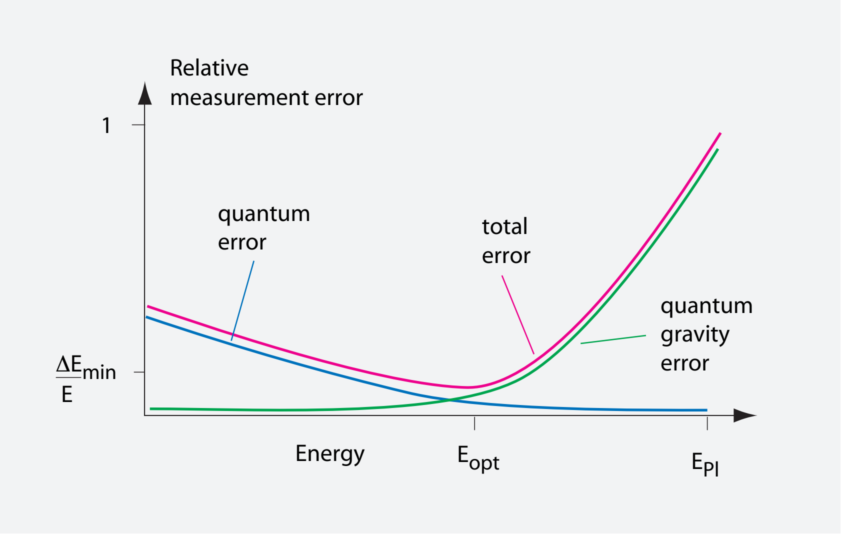


FIGURE 6 Measurement errors as a function of measurement energy.

cases. For *low* energies, the error is due to quantum effects and is given by

$$\frac{\Delta t}{T} \sim \frac{1}{E_{\text{meas}}} \tag{103}$$

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

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

so that the total error varies as shown in Figure 6. 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 7. 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

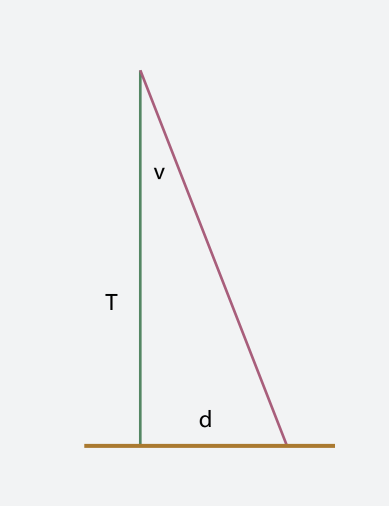


FIGURE 7 Trees and galaxies.

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

$$T = \frac{d}{v} \, . \tag{105}$$

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

$$\frac{\Delta T}{T} = \frac{\Delta d}{d} + \frac{\Delta v}{v} \, . \tag{106}$$

Let us 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 (106), 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 $(13.8 \pm 0.1) \cdot 10^9$ 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 (102) implies

$$\frac{\Delta d}{T} > \left(\frac{l_{\mathrm{Pl}}}{d}\right)^{2/3} \tag{107}$$

---

* 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 8.

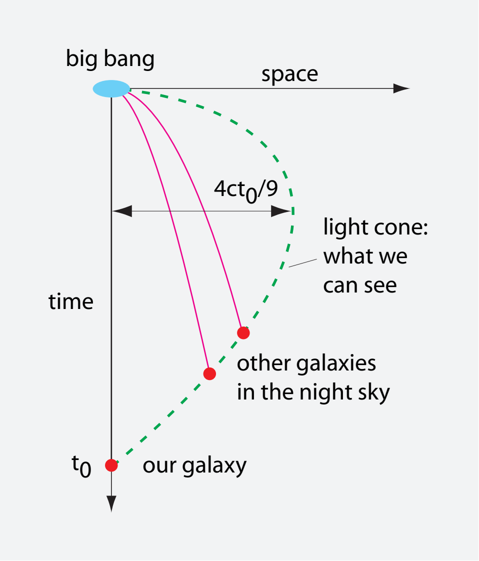


FIGURE 8 The speed and distance of remote galaxies.

thus giving the same indeterminacy in the age of the universe as the one we found above in the case of material clocks. Challenge 58 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. Challenge 59 e 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 (106) 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.*

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 7 to be replaced by the more realistic diagram of Figure 8. The 'light onion' replaces the familiar light cone of special relativity: Ref. 99 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, both the emitted radiation and the object itself are 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 13.8 ± 13.8 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 60 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 on this path is the measurement of length.

MAXIMUM LENGTH

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. Vol. II, page 242

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?

Admittedly, we would have no time to push the metre rule out up to the horizon, because in the standard 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 8 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. Ref. 99 Ref. 99

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. 100 Vol. II, page 307 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

---

[*] 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. Ref. 99

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

$$R(t) = C\, t^{2/3} , \tag{108}$$

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 61 e

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. And again, the size of the universe turns

Challenge 62 e

out not to be a meaningful property.

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 have seen 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, nor 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, measurement errors prevent the determination of the topology of the sky. In fact, this is not a new result. As is well known, general relativity is unable to describe particle–antiparticle pair creation particles with spin 1/2. The reason for this inability is the change in space-time topology required by such processes. The universe is full of these and many other quantum processes; they imply that it is impossible to determine or define the microscopic topology for the universe and, in particular, for the horizon. Can you find at least two other arguments to confirm this conclusion? Challenge 64 s

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 at horizons. Page 58

Finally, there is no way to decide by measurement whether the various points on the horizon are *different* from each other. On the horizon, measurement errors are of the same order as the size of the horizon. The distance between two points in the night sky is thus undefined. Therefore it is unclear what the *diameter* of the horizon is.

---

* The measurement errors also imply that we cannot say anything about translational symmetry at cosmological scales. Can you confirm this? Challenge 63 ny In addition, at the horizon it is impossible to distinguish between space-like and time-like 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 thus confirm the point made above. Page 85

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

Does the universe have initial conditions?

Ref. 101 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 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 58 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.)

Page 303 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. The lack of initial conditions means that we have solved the first issue from the millennium list. Page 19

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? 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. What happens if we combine the two theories? Vol. II, page 307

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 Vol. IV, page 113 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.

In the case of the universe, a comparison with the vacuum is also impossible for purely practical reasons. The particle counter would have to be outside the system. (Can you confirm this?) In addition, it is impossible to remove particles from the universe. Challenge 65 e

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 66 s

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 in the universe, let us weigh the universe.

Does the universe have mass?

Mass distinguishes objects from the vacuum. The average mass 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 Vol. II, page 307

$$\rho_{\mathrm{min}} \approx \frac{m_{\mathrm{min}}(R)}{R^3} = \frac{\hbar}{cR^4} \; . \qquad (109)$$

For the universe, if the standard horizon distance R_0 of 13 800 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 textbook 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 ρ_{min} just mentioned, and the highest possible one, namely the Planck density.* In short, the total mass of the universe depends on the energy of the observer.

Vol. I, page 98 Vol. II, page 64 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 68 e 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 188 average curvature κ for a region of size R, namely

$$\kappa = \frac{1}{r^2_{\text{curvature}}} = \frac{3}{4\pi}\frac{4\pi R^2 - S}{R^4} = \frac{15}{4\pi}\frac{4\pi R^3/3 - V}{R^5} . \qquad (111)$$

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 69 e

Ref. 102 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

$$\Delta\kappa > \frac{16\pi l^2_{\text{Pl}}}{R^4} . \qquad (112)$$

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

We find that *at Planck energy, the average radius of curvature of nature lies between infinity and the Planck length*. This implies that the mass density of the universe lies

---

* In fact, at everyday energies the density of the universe lies midway between the two values, yielding the strange relation Challenge 67 e

$$m_0^2/R_0^2 \approx m_{\text{Pl}}^2/R_{\text{Pl}}^2 = c^4/G^2 . \qquad (110)$$

Vol. V, page 143 But this fascinating relation is not new. The approximate equality can be deduced from equation 16.4.3 (p. 620) of STEVEN WEINBERG, *Gravitation and Cosmology*, Wiley, 1972, namely $Gn_b m_p = 1/t_0^2$. The relation is required by several cosmological models.

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 70 s

In summary, mass measurements of the universe vary with the energy scale. Both at the lowest and at the highest energies, a precise mass value cannot be determined. The concept of mass cannot be applied to the universe as a whole: *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.

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.

Let us explore 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.

Also gauge symmetry is not valid at horizon scale, as you may wish to check in detail yourself. Challenge 71 e 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 scales 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. 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 of the universe.

Challenge 72 s

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.

Ref. 103

We note that if the interior and the boundary of the universe cannot be distinguished, the constituents of nature can neither be points nor tiny objects of any kind. The constituents of nature must be *extended*. But before we explore this topic, we continue with our search for differences between the universe and nothing. The search leads us to our next question.

IS THE UNIVERSE A SET?

> 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 (mathematical) 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

Vol. III, page 322

---

* 'The mistress and queen of all things is reason.' *Tusculanae Disputationes*, 2.21.47. Marcus Tullius Cicero (106–43 BCE), was an influential lawyer, orator and politician at the end of the Roman republic.

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.

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 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, involved mathematical spaces, computer programs, Turing machines, Gödel's incompleteness 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. 104 Challenge 73 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 74 s

In summary, the universe is not a set. In particular, *the universe is not a physical system.* Specifically, it has no state, no intrinsic properties, no wave function, no initial conditions, no energy, no mass, 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. Only parts of nature are approximated or idealized as sets, and thus only parts of nature are physical systems.

---

* 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.

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 or even defined, 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.

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 provokingly 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. 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. *The universe contains no information.* Conclusions such as this one show the power of the criteria that we have deduced for any precise or complete description of motion. Ref. 105

* *

Astronomers regularly 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 75 s

* *

---

* '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.'

** 'Gracious god made the integers, all else is the work of man.' Leopold Kronecker (b. 1823 Liegnitz, d. 1891 Berlin) was a well-known mathematician. Among others, the Kronecker delta and the Kronecker product are named for him.

In 1714, the famous scientist and thinker Leibniz (b. 1646 Leipzig, d. 1716 Hannover) 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. 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?

Ref. 106

Challenge 76 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 77 s

* *

Is the 'radius' of the universe observer-invariant?

Challenge 78 e

* *

Is the cosmological constant Λ observer-invariant?

Challenge 79 e

* *

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

Challenge 80 s

* *

Can you show that the distinction between matter and antimatter is not possible at the cosmic horizon? And the distinction between real and virtual particles?

Challenge 81 e

* *

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 82 s

Hilbert's sixth problem settled

Vol. III, page 276

Ref. 107

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. The problem has remained in the minds of many physicists since that time. Scholars have developed axiomatic treatments if classical mechanics, electrodynamics and special relativity. Then they did this for quantum theory, quantum field theory and general relativity.

Whenever we combine quantum theory and general relativity, we must abandon the concept 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

---

[*] David Hilbert (b. 1862 Königsberg, d. 1943 Göttingen) was the greatest mathematician of his time. His textbooks are still in print.

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 without elements or sets. The reason for this is simple: every (mathematical) *concept* contains at least one element and one set. However, nature is different. 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 and vacuum are defined with the help of objects and objects are defined with the help of space-time and vacuum. In fact, physics has *never* been axiomatic! Physicists have always had to live with circular definitions.

Vol. I, page 433

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. We find that there is air inside clouds, and that 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. In particular, the final, unified theory must contain circular definitions. We will find out how to realize the requirement later on.

Page 217

The perfect physics book

A *perfect* physics book describes *all* of nature with *full precision*. In particular, a perfect physics book describes itself, its own production, its own author, its own readers and its own contents. Can such a book exist?

Since the universe is not a set, a perfect physics book *can* exist, as it does not contradict any property of the universe. Since the universe is not a set and since it contains no information, the paradox of the perfect physics book disappears. Indeed, any existing physics book attempts to be perfect. But now a further question arises.

Vol. I, page 432

TABLE 3 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.

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 247

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 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 83 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.* The term universe has *no* content.

By the way, 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 84 s

---

* 'Thus I have devoted myself to magic, [...] that I understand how the innermost world is held together.' Goethe was a German poet.

Vol. III, page 278

In short, the term ‘universe’ is not at all useful for the description of motion. 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. And the same applies to the term ‘nature’. 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. In short, since the term ‘universe’ has no content, the term ‘multiverse’ has even less. The latter term has been created only to trick the media and various funding agencies. In fact, the same might be said of the former term...

Challenge 85 e

So far, 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. *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 19

At the beginning of this volume, 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 compare 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.

---

* 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 scales
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	do not apply	apply	do 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 and gauge symmetries	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

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!

Secondly, and more importantly, nature behaves in the *same way* at the cosmological horizon scale and at the Planck scale. 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 86 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.

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.

Challenge 87 e

Ref. 108

The simplest approach to extremal identity – in fact, one that is too simple to be correct – is *inversion*. Indeed, extremal identity seems to imply a connection such as

$$r \leftrightarrow \frac{l_{\mathrm{Pl}}^2}{r} \quad \text{or} \quad x_\mu \leftrightarrow \frac{l_{\mathrm{Pl}}^2\, x_\mu}{x_\mu\, x^\mu} \tag{113}$$

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 (113) 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?

Challenge 88 s

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

Ref. 103

Above all, we need to find the correct version of the inversion relation (113). 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.

Challenge 89 e

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.

Page 85

SUMMARY ON THE UNIVERSE

The exploration of the universe allows us to formulate several additional requirements for the final theory that 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 meaningless whenever *full precision* is required.
- We learned that a description of nature without sets *solves the contradictions* between

general relativity and quantum theory.
- We found, again, that despite the contradictions between quantum theory and general relativity, the Planck limits c, $\hbar$ and $c^4/4G$ *remain valid*.
- 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 close relationship between large and small scales in nature.

We can now answer the question in the chapter title: there seems to be little difference – if any at all – between the universe and nothing. We can express this result in the following catchy statement: the universe cannot be observed. The confusion and tension are increasing. But in fact we are getting close to our goal, and it is worth continuing.

A PHYSICAL APHORISM

Ref. 55, Ref. 56

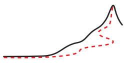
Here is a humorous 'proof' that we really are near the top of Motion Mountain. Salecker and Wigner, and then Zimmerman, formulated the fundamental limit for the measurement precision τ attainable by a clock of mass M. It is given by $\tau = \sqrt{\hbar T/c^2 M}$, where T is the time to be measured. We can then ask what maximum 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

$$T = \frac{t_{\mathrm{Pl}}^2 c^2}{\hbar} M \,. \tag{114}$$

Challenge 90 s

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 scales. We are thus getting nearer to the top of Motion Mountain. Be prepared for a lot of fun.



Chapter 6

THE SHAPE OF POINTS – EXTENSION IN NATURE

> Nil tam difficile est, quin quaerendo investigari possiet.[**]
>
> Terence

The usual expressions for the reduced Compton wavelength $ƛ = \hbar/mc$ and for the Schwarzschild radius $r_s = 2Gm/c^2$, taken together, imply 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 result with the following arguments:

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 consists of points, as they cannot be measured or detected. In addition, in order to build up a measurable entity, such as the vacuum, that is extended in three dimensions, its constituents must also 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 (mathematical) set. And 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 extended constituents at high energies, namely as braids or tangles.
7. The belt trick provides a model for fermions that matches observations and again suggests extended constituents in matter.

We conclude the chapter with some experimental and theoretical checks of extension

---

[**] 'Nothing is so difficult that it could not be investigated.' Terence is Publius Terentius Afer (b. *c.* 190 Carthago, d. 159 BCE Greece), important Roman poet. He writes this in his play *Heauton Timorumenos*, verse 675.

and an overview of present research efforts.

> 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.

Page 88

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 to find 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. There must be *uncuttable* particles, which they called *atoms*. Twenty-five centuries later, experiments in the field of quantum physics confirmed the conclusion, but modified it: nowadays, the elementary particles are the 'atoms' of matter and radiation.

---

* '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.' Erwin Schrödinger (b. 1887 Vienna, d. 1961 Vienna) discovered the equation that brought him international fame and the Nobel Prize in Physics.

Despite the huge success of the concept of elementary particle, at Planck energy, we have a different situation. The use of a knife, like any other cutting process, is the insertion of a wall. Walls and knives are potential hills. All potential hills are of finite height, and allow tunnelling. Therefore a wall is never perfect, and thus neither is a knife. In short, any attempt to divide matter fails to work when we approach Planck scales. At Planck energy, any subdivision is impossible.

The limitations of knives and walls imply that at Planck energy, an attempted cut does not necessarily lead to two separate parts. At Planck energy, 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, at Planck scales we cannot prove compactness by cutting objects.

Are cross sections finite?

To sum up: despite all attempts, we cannot show that elementary particles are point-like. 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. Ref. 109 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 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.

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 beyond all bounds. So at high energies we cannot give an upper limit to the size of a particle – except the universe itself. In other words, since particles are not point-like at everyday energies, at Planck energy they are enormous:

▷ Quantum particles are extended.

That is quite a statement. Are particles really not of finite, bounded size? Right at the Vol. I, page 24 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. But all these basic assumptions fail us at Planck scales. At Planck energy, it is impossible to determine whether something is bounded or compact. Compactness and locality 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.

TABLE 5 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

We conclude that particles at Planck scales are as extended as the vacuum. Let us perform another check.

CAN WE TAKE A PHOTOGRAPH OF A POINT?

> “Καιρὸν γνῶθι.*”
>
> Pittacus

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. 110

Ref. 47, Ref. 24

Imagine that you have the world's best shutter and that you are taking photographs at shorter and shorter times. Table 5 gives a rough overview of the possibilities. When shutter times are shortened, photographs get darker and sharper. 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

---

* 'Recognize the right moment.' Also rendered as: 'Recognize thy 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.

more *blurred* and *incorrect* 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.

In short, the blur decreases when shutter times usual in everyday life are shortened, but *increases* when shutter times are shortened further 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. In fact, a shutter does not exist even at medium energy: shutters, like walls, stop existing at around 10 MeV. 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 need to study its 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. 47

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. 47 Challenge 91 e

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 thought 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. Page 85

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 106

However, an elementary particle *can* have constituents, provided that they are not compact. The difficulties of compact constituents were described by Andrei 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. 28

SUMMARY OF THE FIRST ARGUMENT FOR EXTENSION

Point particles do not exist at Planck scales. At Planck scales, all thought experiments with partciles suggest that matter and radiation are made of *extended and fluctuating constituents of infinite size.*

Ref. 47 For extended constituents, the requirement of a non-local description 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 these constituents are extended and massless. However, if the constituents are infinitely extended, how can compact, point-like particles be formed from them? We will look at a few options shortly. Challenge 92 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. 112

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

---

[*] 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.

[**] Thus at Planck scales there is no quantum Zeno effect. Ref. 111

[***] Aristotle (b. 384/3 Stageira, d. 322 BCE Chalkis), Greek philosopher and scientist.

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.

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 Ref. 47 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. And in nature, an intrinsic length exists for matter and for vacuum: the Planck length. And 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.

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 result 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

---

* Imagining the vacuum as a collection of compact constituents, such as spheres, with Planck size in all Ref. 113 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. Challenge 93 s

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 thought 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.

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 and the constituents of space 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

Planck scales imply that *space is made of fluctuating extended constituents of huge size.* Like particles, also space and vacuum are not made of points, but of a web. Vacuum requires a statistical description.

Vol. I, page 335 More than two thousand years ago, the Greeks argued that matter must be made of Ref. 114 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 (partially) 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 research 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 usually conjectured to have a value somewhere between 1 and 1000. Ref. 115

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. 116

---

[*] William Shakespeare (1564 Stratford upon Avon–1616 Stratford upon Avon) wrote theatre plays that are treasures of world literature.

[**] 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.

[***] Simplicius of Cilicia (*c.* 499 – 560), neoplatonist philosopher.

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 $l_{\rm Pl}^2/R$ cannot be distinguished. Lengths make no sense when they are smaller than $l_{\rm Pl}$. 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 $R \leftrightarrow l_{\rm Pl}^2/R$ 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 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

So far, 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. Vol. III, page 89 It includes, among others, the famous electric–magnetic duality we first encountered in electrodynamics, Ref. 117 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. Ref. 115

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

---

[*] A symmetry between size and Schwarzschild radius, i.e., a symmetry between length and mass, leads 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 implies a symmetry between coordinates and wave functions, i.e., a symmetry between states and observables. It leads to quantum theory.

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.*

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 94 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 95 e

---

* Renormalization energy does connect different energies, but not in the correct way; in particular, it does not include duality.

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. 118

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. 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 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 fails to explain 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 322

---

[*] ‘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 Munich), 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: the existence of *any* multitude in nature is questionable.

Ref. 119

> 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.

In modern physics, many specialized sets are used to describe nature. We have used vector spaces, linear spaces, topological spaces and Hilbert spaces. But so far, 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.

Page 101

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.

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. Circularity is not a good idea, but we can live with it – at low energy. But at Planck energy, vacuum and particles are indistinguishable from each other. Particles and vacuum – thus everything – are the same. We have to abandon the circular definition. This is a satisfactory outcome; however, it also implies that nature is not a set.

Also space-time duality implies that space is not a set. Space-time duality 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 impossibility to use a set to describe space-time.

Ref. 121

In short,

---

* 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 and Zeno. (Their home city, Elea, was near Naples.) Plato does not reach a conclusion. Modern physics, however, does.

Ref. 120

▷ 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 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 285

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, reaches spatial infinity, allows approximate localization, and thus allows 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 we try 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 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 below. Page 146

Challenge 96 r * This is the simplest model; but is it the only way to describe nature?

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

$$r_S = 2Gm/c^2 \tag{115}$$

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. 57, Ref. 58

$$S = \frac{k}{4l_{\text{Pl}}^2} A = \frac{kc^3}{4\hbar G} A \quad \text{or} \quad S = k\frac{4\pi G m^2}{\hbar c} \tag{116}$$

where k is the Boltzmann constant and $A = 4\pi r_S^2$ is the surface of the black hole horizon. Ref. 122 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. 35 not only for black holes. Jacob 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 Ref. 123 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 Ref. 125 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

---

Ref. 124 * The result can be derived from quantum statistics alone. However, this derivation does not yield the proportionality coefficient.

Ref. 126 kept a distance r apart, is given by

$$S(r) = k\frac{3r^2}{2Na^2} \quad \text{for} \quad Na \gg \sqrt{Na} \gg r\,. \tag{117}$$

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 (117) 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, we get 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 $a\sqrt{N} \approx l_{\text{Pl}}$; 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 Ref. 109 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. 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. But if a system is defined by its surface, as a black hole is, its components must be extended.

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 long as all matter and radiation particles are made of the same extended components. The final state of a black hole only depends on the number of extended constituents.

SUMMARY OF THE FIFTH ARGUMENT FOR EXTENSION

Black hole entropy is best understood as resulting from extended constituents that tangle Page 285 and fluctuate. And black hole entropy 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

$$[x, y] = xy - yx = 0 \qquad (118)$$

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

$$[x, y] = l_{\mathrm{Pl}}^2 + ... \; . \qquad (119)$$

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, Ref. 47 gravity implies that a double exchange does not lead back to the original situation at Planck scales.

Constituents obeying this or similar relations have been studied in mathematics for Ref. 127 many decades: they are called *braids*. Thus space is not made of points at Planck scales, but of braids or their generalizations, namely tangles. We find again that quantum theory and general relativity taken together imply that the vacuum must be made of extended constituents.

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

$$a^{\dagger}b^{\dagger} \pm b^{\dagger}a^{\dagger} = 0 \; . \qquad (120)$$

At Planck energy this cannot be correct. Quantum gravity effects modify the right-hand side: they add an energy-dependent term that is negligible at experimentally accessible energies but becomes important at Planck energy. We know from our experience with

---

* The same reasoning applies to the so-called fermionic or Grassmann coordinates used in supersymmetry. They cannot exist at Planck energy.

Ref. 47 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 (120) for which the right-hand side does not vanish is *braid* Ref. 127 *symmetry*. This again suggests that particles are made of extended constituents.

Summary of the sixth argument for extension

Extrapolating both point and particle indistinguishability to Planck scales suggests extended, braided or tangled 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 detection of exchange.* In other words, any part of the universe is also described by its *exchange behaviour*.

In nature, particle 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, Vol. IV, page 110 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, like one of those shown in Figure 9, 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 this point. The tail model even has the same problems with highly curved Page 171 space as real spin $1/2$ particles have. We will explore the issues in more detail shortly.

---

[*] 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.

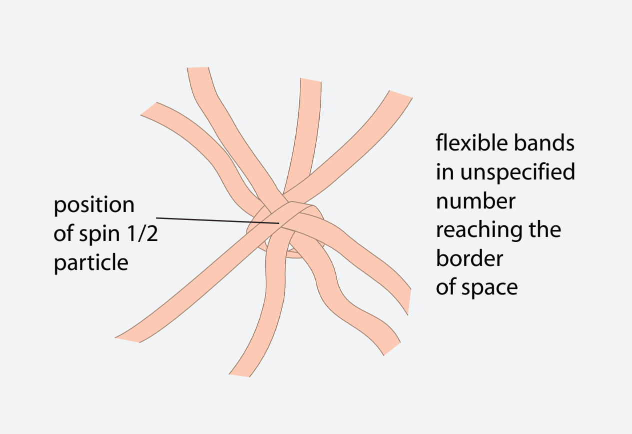


FIGURE 9 Possible models for a spin 1/2 particle.

The tail model thus confirms that rotation is partial exchange. More interestingly, it shows that rotation implies connection with the border of space. Extended particles can be rotating. Particles can have spin 1/2 provided that they have tails going to the border of space. If the tails do not reach the border, the model does not work. Spin 1/2 thus even seems to *require* extension. Ref. 128

Challenge 97 e It is not hard to extend this idea to include spin 1 particles. In short, both bosons and fermions can be modelled with extended constituents.

SUMMARY OF THE SEVENTH ARGUMENT FOR EXTENSION

Exploring the properties of particle spin suggests the existence of extended constituents in elementary fermions. We note that gravitation is not used explicitly in the argument. It is used implicitly, however, in the definition of the locally flat space-time and of the asymptotic region to where the tails are reaching.

CURIOSITIES AND FUN CHALLENGES ABOUT EXTENSION

> “No problem is so big or complicated that it can’t be run away from.”
>
> Charles Schulz

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 98 s

* *

Challenge 99 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 100 s

* *

Does duality imply that a system with two small masses colliding is equivalent to a system with two large masses gravitating? Challenge 101 s

* *

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 102 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 103 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 104 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 105 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 106 d

* *

In 2002, Andrea Gregori made a startling 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. Is this conclusion unavoidable? Ref. 129 Challenge 107 s

* *

What is wrong with the following argument? We need lines to determine areas, and we need areas to determine lines. This implies that at Planck scales, we cannot distinguish

Challenge 108 s areas from lengths at Planck scales.

* *

Ref. 129, Ref. 130 Challenge 109 s We need a description for the expansion of the universe in terms of extended constituents. Various approaches are being explored. Can you speculate about the solution?

Gender 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 description of nature as made of tiny balls moving in a void, we called this as a typically male idea. This implies that the female part is missing. Which part would that be? Vol. I, page 335

From a general point of view, the female part of physics might be the quantum description of the vacuum, the container of all things. 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. And women might not have needed 2500 years to understand that nature is not made of a void and little balls, but that everything in nature is made of extended constituents. It is curious that (male) physics took so long for this discovery.

CHECKS OF EXTENSION

The idea that nature is described by extended constituents is taken for granted in all current research approaches to unification. How can we be sure that extension is correct? The arguments presented so far provide several possible checks. We start with some options for *theoretical* falsification.

- Any explanation of black hole entropy *without* extended constituents would invalidate the need for extended constituents.
- A single thought experiment *invalidating* extended constituents would prove extension wrong. Challenge 110 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, or duality, would invalidate the need for extended constituents. As Edward Witten likes to say, any unified model of nature must include duality. Page 111
- If the measurement of length could be shown to be *unrelated* to the counting of folds of extended constituents, extension would become unnecessary.
- Finding any property of nature that *contradicts* extended constituents would spell the end of extension.

Any of these options would signal the end for almost all current unification attempts. Fortunately, theoretical falsification has not yet occurred. 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. 131
- Paul Mende has proposed a number of checks on the motion of extended objects in space-time. He argues that an extended object and a mass point move differently; the differences could be noticeable in scattering or dispersion of light near masses. Ref. 132

Experimental falsification of extension has not yet occurred. In fact, experimental falsification is rather 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. Page 19 Among others, confirmation requires to find a concrete model, based on extended constituents, for the electron, the muon, the tau, the neutrinos, the quarks and all bosons. Confirmation also requires using extended constituents to realize an old dream of particle physics: to deduce the values of the coupling constants and particle masses. Before we attempt this deduction, we have a look at some other attempts.

Current research based on extended constituents

> “To understand is to perceive patterns.”
> Isaiah Berlin*

The Greeks deduced the existence of atoms from the observation that fish can swim through water. Ref. 114 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 extended constituents.

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. Ref. 133 It also implies coupling duality, and thus the extension of fundamental constituents.

* *

Ref. 134, Ref. 135 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.

* *

---

* Isaiah Berlin (b. 1909 Riga, d. 1997 Oxford) was an influential political philosopher and historian of ideas.

Ref. 136 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. 137 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 155 It is long known that wave function collapse can be seen as the result of extended constituents. We will explore the details below.

* *

Ref. 138, Ref. 139, Ref. 140, Ref. 141, Ref. 142, Ref. 143, Page 345 At the start of the twenty-first century, a number of new 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. Several of them are discussed in more detail below.

Despite the use of extension, none of these attempts solved a single problem from the millennium list. One approach – especially popular between the years 1984 and 2010 – merits a closer look.

SUPERSTRINGS – EXTENSION PLUS A WEB OF DUALITIES

> “Throw physic to the dogs; I'll none of it.
> William Shakespeare, *Macbeth*.”

Ref. 144 *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 superstrings*. In a subsequent generalization, particles are conjectured to be fluctuating higher-dimensional *supermembranes*.
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 superstrings 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. 145 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 *superstring conjecture* – named so by Brian Greene, one of its most important researchers – acquires a number of appealing characteristics. Ref. 146

WHY SUPERSTRINGS AND SUPERMEMBRANES ARE SO APPEALING

First of all, the superstring 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. (However, sets are still used.)

The superstring 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 of other inconsistencies. Historically, the first example was the Green–Schwarz anomaly cancellation; superstrings also solve other anomalies and certain inconsistencies 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, superstrings cure the infinities that appear in m^2/r when the distance r goes to zero.*

Also following Witten, in the superstring conjecture, the interactions follow from the particle definitions: interactions do not have to be added. That is why the superstring conjecture predicts gravity, gauge theory, supersymmetry and supergravity.

About gravity, one of the pretty results of the superstring conjecture is that superstrings and black holes are complementary to each other. This was argued by Polchinsky, Horowitz and Susskind. As expected, superstrings explain the entropy of black holes. Strominger and Vafa showed this important result in 1996. Ref. 147 Ref. 122

The superstring 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 the superstring conjecture in certain higher-dimensional spaces that contain anti-de Sitter space.

In short, the superstring 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.

---

* This argument is questionable, because general relativity already cures that divergence.

WHY THE MATHEMATICS OF SUPERSTRINGS IS SO DIFFICULT

The superstring 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 superstring of matter. Then the Lagrangian for the corresponding quantum superstring fields is constructed, and then higher dimensions, supersymmetry, dualities and membranes are incorporated. This formulation of the superstring conjecture takes for granted the existence of a space-time background.

The Lagrangian of the superstring 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 superstring 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 superstring 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.

Indeed, a few years ago a physicist and a mathematician listened to a talk on superstrings, describing nature in eleven dimensions. The mathematician listened intensely and obviously enjoyed the talk. The physicist did not understand anything and got more and more annoyed. At the end, the physicist had a terrible headache, whereas the mathematician was full of praise. ‘But how can you even understand this stuff?’, asked the physicist. ‘I simply picture it in my head!’ ‘But *how* do you imagine things in eleven dimensions?’ ‘Easy! I first imagine them in N dimensions and then let N go to 11.’

TESTING SUPERSTRINGS: 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. 148

$$m_{\text{proton}} \sim \mathrm{e}^{-k/\alpha_{\text{Pl}}}\, m_{\text{Pl}} \quad \text{and} \quad k = 11/2\pi\,,\ \alpha_{\text{Pl}} \approx 1/25\,. \tag{121}$$

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 $\alpha_{\text{Pl}} \approx 1/25$ is an empirical value based on experimental data. Page 376

Any unified theory must allow us to calculate the three gauge coupling constants as a function of energy, thus also α_{Pl}. 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 superstring conjecture. Without knowledge of the vacuum state, no calculations of coupling constants or masses are possible.

The vacuum state of the superstring conjecture is expected to be one of an rather involved set of topologically distinct manifolds. Ref. 149 It is estimated that there are between 10^{500} and $10^{272\,000}$ candidate vacuum states. The universe contains 10^{80} atoms; it thus seems much easier to find a particular atom in the universe than to find the correct vacuum state. The advantages that are due to a unique Lagrangian are thus lost.

We can also describe the problems with the calculation of particle masses in the following way. The superstring 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 numerical properties have not yet been deduced from superstrings.

The status of the superstring conjecture

> “Es ist nichts Großes ohne Leidenschaft vollbracht worden, noch kann es ohne solche vollbracht werden.*”
>
> Friedrich Hegel, *Enzyklopädie*.

It is fair to say that nowadays, superstring researchers are stuck. Despite the huge collective effort, not a single calculation of an *experimentally measurable* value has been performed. For example, the superstring conjecture has not predicted the masses of any elementary particle, nor the value of any coupling constant, nor the number of gauge interactions, nor the number of particle generations. In fact, *none* of the open issues in physics that are listed the millennium list has been solved by the superstring conjecture. Page 19 This disappointing situation is the reason that many scholars, including several Nobel Prize winners, dismiss the superstring conjecture altogether. Ref. 150

What are the reasons that the superstring conjecture, like several other approaches based on extended constituents, was unsuccessful? Superstrings and supermembranes are *complex*: superstrings and supermembranes move in many dimensions, carry mass, have tension and carry (supersymmetric) fields. In fact, the precise mathematical definition of a superstring or a supermembrane 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.

Superstrings are not simple entities. And no researcher tried to make them simple. Put in different terms, the superstring conjecture was not successful because its *basic principles* have never been clarified. Ref. 151 It is estimated that, from 1984 to 2010, over 10 000 man-years have been invested in the exploration of the superstring conjecture. Compare this with about a dozen man-years for the foundations and principles of electrodynamics, a dozen man-years for the foundations and principles of general relativity, and a dozen man-years for the foundation and principles of quantum theory. The lack of clear foundations of the superstring conjecture is regularly underlined even by its supporters, such as Murray Gell-Mann. And despite this gap, *no* research papers on the basic principles exist – to this day.

There is a second reason for the lack of success of the superstring conjecture: superstrings and supermembranes are based on sets. And we saw above that no unified theory can be based on sets.

---

* ‘Nothing great has been achieved without passion, nor can it be achieved without it.’ Hegel, an influential philospher, writes this towards the end of the third and last part of his *Enzyklopädie der philosophischen Wissenschaften im Grundrisse*, §474, 296.

SUMMARY ON EXTENSION IN NATURE

> “Wir müssen wissen, wir werden wissen.*”
> David Hilbert

We have explored nature at her limits: we have studied the Planck limits, explored three-dimensionality, curvature, particle shape, renormalization, spin and bosonization; we have investigated the cosmological constant problem and searched for a ‘background-free’ description of nature. As a result, we have found that at Planck scales, all these explorations lead to the same conclusions: Ref. 152

- Points and sets do not describe nature correctly.
- Matter and vacuum are two sides of the same medal.
- What we usually call space-time points and point particles are in fact made up of common and, above all, *extended* constituents.

We can reach the conclusions in an even simpler way. What do quantum theory and black holes have in common? They both suggest that nature is made of extended entities. We will confirm below that both the Dirac equation and black hole entropy imply that particles, space and horizons are built from extended constituents.

Despite using extension as fundamental aspect, and despite many interesting results, *all* the attempts from the twentieth century – including the superstring conjecture and all quantum gravity models, but even supersymmetry and supergravity – have *not been successful* in understanding or in describing nature at the Planck scale. The reasons for this lack of success were the unclear relation to the Planck scale, the lack of clear principles, the use of sets or other incorrect assumptions, and, above all, the unclear connection to experiment. At the latest in 2014, during the Strings conference, it became clear that the string research community has quietly given up its quest to achieve a unified theory with the help of superstrings or supermembranes. Researchers are now looking for other microscopic models of nature.

To be successful, we need a different approach to calculations with extended constituents. We need an approach that is built on Planck units, is based on clear principles, has few but correct assumptions, and provides predictions that stand up against experimental tests.

In our quest for a final theory of physics, one way to advance is by raising the following issue. The basis for the most explored approach in the late twentieth century, the superstring conjecture, is formed by four 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 wavelength imply, as we found out earlier on, that the dimensionality of space and the statistics of particles are *undefined* at Planck scales. In other words, nature does not have higher dimensions or supersymmetry at Planck scales. And all known experiments confirm this conclusion. In our quest for a final theory of motion, we therefore drop the two incorrect assumptions and continue our adventure. Page 71, page 77 Page 74, page 77

We can phrase the remaining quest in the following specific way:

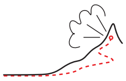
---

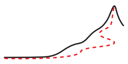
* ‘We must know, we will know.’ This was Hilbert’s famous personal credo.

- ▷ How are the three gauge interactions related to the fundamental constants c, $\hbar$, k and G, i.e., to the Planck scale?

Challenge 111 e

This question is rarely asked. Attempts to answer it are even rarer. (Can you find one?) Finding the Planck origin of gauge interactions means finding the final theory. To be successful in this quest, we need two tools: extension and simplicity.





CHAPTER 7

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 motion that were discovered in the twentieth century – quantum field theory and general relativity – are the low-energy approximations of how nature behaves at Planck scales. In order to understand nature at Planck scales, and thus to find the unified and final description of motion, 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 a promising speculation for the unified and final theory of motion.

REQUIREMENTS FOR A FINAL THEORY

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 detailed below, the unified description must be valid for *all observations* and provide *complete precision*. Page 161

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 of all, only constituents that are extended allow us to deduce black hole entropy. Secondly, only extended constituents allow us to model that elementary particles are not point-like or that physical space is not made of points. Thirdly, only extended constituents allow us to model a smallest measurable space and time interval. Fourthly, only extended constituents allow us to model spin 1/2 in locally flat space-time.

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 show below, if a candidate for a final theory can be modified, or generalized, or reduced to special cases, or varied in any other way, it is not final. Page 163

In the preceding chapters we have deduced many additional, requirements that a final

---

[**] Ernest Rutherford (b. 1871 Brightwater, d. 1937 Cambridge) was an important physicist and researcher; he won the Nobel Prize in Chemistry for his work on atoms and radioactivity.

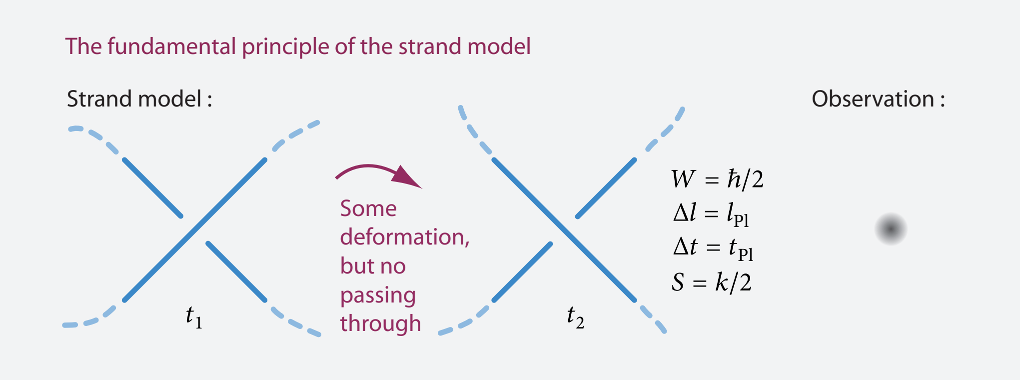


FIGURE 10 The fundamental principle of the strand model: the simplest observation in nature, a ''point-like'' fundamental event, is defined by a crossing switch in three spatial dimensions. The crossing switch defines the action $\hbar/2$, the Planck length, the Planck time and half the Boltzmann constant $k/2$.

theory must realize. The full list of requirements is given in Table 6. Certain requirements follow from 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 with quantum theory and general relativity. More specifically, every requirement appears when the expressions for the Compton wavelength and for the Schwarzschild radius are combined. So far, the table is not found elsewhere in the research literature.

Challenge 112 e

TABLE 6 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 precisely describe all motion – everyday, quantum and relativistic – and explain all open issues from the millennium list, given (again) in Table 8 on page 161, including the fine structure constant.
Modification	must be *impossible*, as explained on page 163.
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 therefore must be described by *common* fundamental constituents.
Fundamental constituents	must determine all observables.
Fundamental constituents	must be as *simple* as possible, to satisfy Occam's razor.
Fundamental constituents	must be *extended* and *fluctuating*, to explain black hole entropy, spin, minimum measurement intervals, space-time homogeneity and isotropy of space.
Fundamental constituents	must be the *only unobservable* entities. (If they were observable, the theory would not be final, because the properties of the entities would need explanation; if additional unobservable entities would exist, the theory would be fiction, not science.)

TABLE 6 (Continued) General requirements for a final and unified description of nature and of motion.

ASPECT	REQUIREMENTS FOR THE FINAL AND UNIFIED DESCRIPTION
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* at Planck scales, 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.
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.
Singularities	must not exist, due to the limits of measurements.
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.
Quantum field theory, including QED, QAD, QCD	must follow from the final unified theory by eliminating G.
General relativity	must follow from the final unified theory by eliminating $\hbar$.
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 scales.
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 scales, as space is undefined there.
Symmetries	must be *undefined* at Planck scales, due to the limits to measurement precision.
Large and small scales	must be *similar*, due to the limits to measurement precision.

Challenge 113 e

The requirement list given in Table 6 can be considerably shortened. Shortening the list of requirements is possible because the various requirements are *consistent* with each other. In fact, shortening is possible because a detailed check confirms a suspicion that arose during the last chapters: extension alone is sufficient to explain all those requirements that seem particularly surprising or unusual, such as the lack of points or the lack of axioms. Such a shortened list also satisfies our drive for simplicity. After shortening, two requirements for a final theory remain:

▷ *The final theory must describe nature* at and below the Planck scale[*] *as made of extended constituents fluctuating in a background.* Extended constituents must explain particles, space and horizons.

▷ In the final theory, the *fluctuations* of the extended constituents *must explain all motion.* The Planck-scale fluctuations must describe all observed examples of everyday, quantum and relativistic motion with complete precision, imply all concepts of physics and explain all fundamental constants.

This requirement summary is a result of our adventure so far. It forms the starting point for the final leg of our adventure. If you do not agree with these two requirements, take a rest and explore your disagreement in all its details. Challenge 114 s

Looking at the table of requirements for the final theory – both the full one and the shortened one – 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! This unexpected conclusion is the main reason that past attempts failed to discover the final theory. In fact, most attempts do not fulfil the requirements because various scholars explicitly disagree with one or several of them.

The requirement of the *extension* of the fundamental constituents is the main result about the properties of a final theory. A final theory must make a statement about these constituents. The fundamental constituents, also called *fundamental degrees of freedom*, must explain everything we observe and know about nature. In particular, the constituents must explain the curvature of space, the entropy of black holes and the spectrum, mass and other properties of all elementary particles. And these constituents must be extended. Extension is the reason that the final theory contradicts both general relativity and quantum theory – but still allows these theories as excellent approximations. In short, extension is the key to finding the final theory.

The requirement summary implying fluctuating extended constituents resulted from our drive for extreme simplicity. With this summary, the search for a candidate final theory does not take long. Of the few candidates that satisfy the requirements for a final theory, it seems that the *simplest* is the one based on *fluctuating featureless strands*. In this approach, *strands*,[**] not points, are assumed to be the fundamental constituents of vacuum, horizons, matter and radiation. Ref. 153 Page 165

INTRODUCING STRANDS

The strand model starts with a simple idea:

▷ Nature is made of unobservable, fluctuating, featureless strands.

---

[*] The final theory must not describe nature *beyond* the Planck scales. A more relaxed requirement is that the predictions of the theory must be *independent* of any fantasies of what might occur *beyond* Planck scales.

[**] In Dutch: draden, in French: fils, in German: Fäden, in Italian: fili.

We will discover that everything observed in nature – vacuum, fermions, bosons and horizons – is made of strands. Strands are the common and extended constituents of everything. Even though strands are unobservable and featureless, all observations are due to strands.

▷ All observations, all change and all events are composed of the fundamental event, the crossing switch.

To describe all observations with precision, the strand model uses only one *fundamental principle*:

▷ ***Planck units*** are defined through **crossing switches** of strands.

Page 147 The definition of the Planck units with the crossing switch is illustrated in Figure 10. All measurements are consequence of this definition. All observations and everything that happens are composed of fundamental events. The fundamental principle thus specifies why and how Planck units are the *natural units* of nature. In particular, the four basic Planck units are associated in the following way:

▷ ***Planck's quantum of action*** $\hbar/2$ appears as the action value associated to a crossing switch. The action $\hbar$ corresponds to a double crossing switch, or full turn of one strand segment around another.*

▷ The (corrected) ***Planck length*** $l_{\mathrm{Pl}} = \sqrt{4G\hbar/c^3}$ appears as the effective diameter of strands. Since the Planck length is a limit that cannot be achieved by measurements, strands with such a diameter remain unobservable.*

▷ The ***Planck entropy***, i.e., the Boltzmann constant k, is the natural unit associated to the counting and statistics of crossings.*

▷ The (corrected) ***Planck time*** $t_{\mathrm{Pl}} = \sqrt{4G\hbar/c^5}$ appears as the shortest possible duration of a crossing switch.*

Crossing switches that are faster than the Planck time do not play a role, as they are unobservable and unmeasurable. Let us see why.

How can we imagine a minimum time interval in nature? A crossing switch could be arbitrarily fast, couldn't it? So how does the Planck time arise? To answer, we must recall the role of the observer. The observer is a physical system, also made of strands. The observer cannot define a really continuous background space-time; careful consideration tells us that the space-time defined by the observer is somewhat fuzzy: it is effectively *shivering*. The average shivering amplitude is, in the best possible case, of the order of a Planck time and length. Therefore, crossing switches faster than the Planck time are not observable by an observer made of strands.**

---

* In other words, the strand model sets $\hbar = l_{\mathrm{Pl}} = t_{\mathrm{Pl}} = k = 1$. The strange numbers that these constants have in the SI, the international system of units, then follow automatically from the definitions of the metre, second, kilogram and kelvin.

** The issue of time remains subtle also in the strand model. The requirement of consistency with macro-

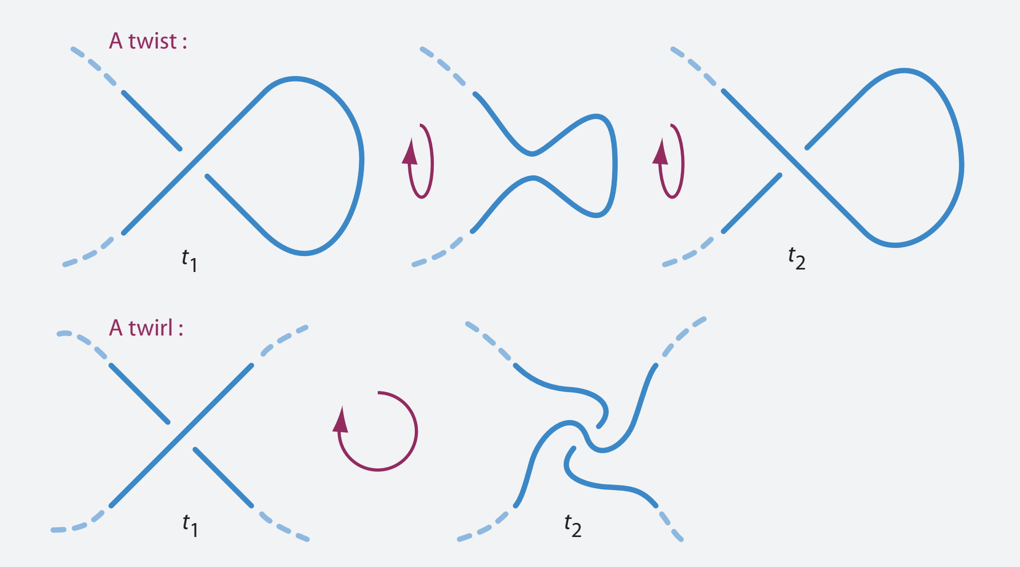


FIGURE 11 An example of strand deformation leading to a crossing switch (above) and one that does *not* lead to a crossing switch (below).

Strands are *impenetrable*; the switch of a crossing thus always requires the motion of strand segments *around* each other. The simplest example of a deformation leading to a crossing switch is shown in Figure 11.

Can you deduce the strand processes for the Planck momentum, the Planck force and the Planck energy? Challenge 116 s

Challenge 117 e Exploring strand processes we find: the fundamental principle implies that every Planck unit is an observer-invariant *limit value*. Therefore, the fundamental principle naturally contains special and general relativity, quantum theory and thermodynamics (though not elementary particle physics!). Page 36 In theory, this argument is sufficient to show that the fundamental principle contains all these parts of twentieth century physics. In practice, however, physicists do not change their thinking habits that quickly; thus we need to show this result in more detail.

EVENTS, PROCESSES, INTERACTIONS AND COLOURS

In the strand model, every physical process is described as a sequence of crossing switches. But every physical process is also a sequence of events. We thus deduce that events are processes:

▷ Any ***event***, any observation, any measurement and any interaction is composed of switches of crossings between two strand segments.

---

scopic experience, realized with shivering space or space-time, allows us to side-step the issue. An alternative approach might be to picture a crossing switch and its fluctuations in 4 space-time dimensions, thus visualizing how the minimum time interval is related to minimum distance. This might be worth exploring. Challenge 115 r But also in this approach, the fuzziness due to shivering is at the basis of minimum time.

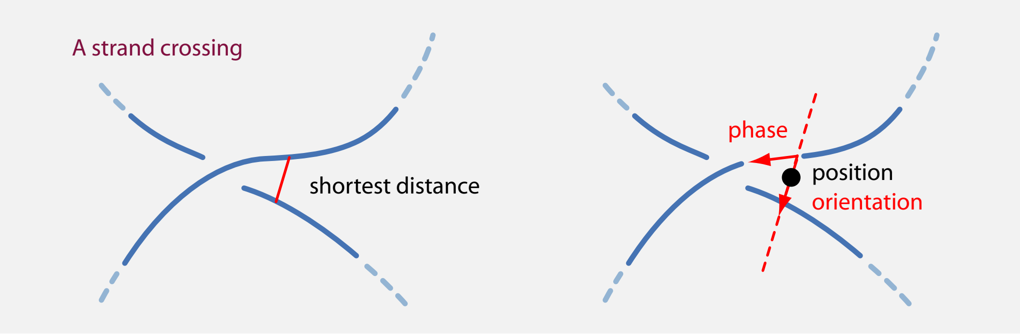


FIGURE 12 The definition of a crossing, its position, its orientation and its phase.

The crossing switch is the fundamental process in nature. We will show that describing events and interactions with the help of crossing switches leads, without alternative, to the *complete* standard model of particle physics, with all its known gauge interactions and all its known particle spectrum.

In particular, we will show:

▷ Particle masses, the ***elementary electric charge*** e and the ***fine structure constant*** $\alpha = 137.036(1)$ are due to crossing switches.

The value of the fine structure constant and the standard model are *not evident* consequences of the fundamental principle; nevertheless, they are *natural* consequences – as we will find out.

From strands to modern physics

Every observation and every process is a sequence of crossing switches of unobservable strands. In turn, crossing switches are automatic consequences of the shape fluctuations of strands. We will show below that all the continuous quantities we are used to – physical space, physical time, gauge fields and wave functions – result from *averaging* crossing switches over the background space. The main conceptual tools necessary in the following are:

▷ A ***crossing*** of strands is a local minimum of strand distance. The position, orientation and phase of a crossing are defined by the space vector corresponding to the local minimum of distance, as shown in Figure 12.

Page 171 The position, orientation and phase of crossings will lead, as shown later on, to the position, orientation and phase of wave functions. The sign of the orientation is defined by arbitrarily selecting one strand as the starting strand. The even larger arbitrariness in the definition of the phase will be of great importance later on: it implies the existence of the three known gauge groups.

▷ A ***crossing switch*** is the rotation of the crossing orientation by an angle π *at* *a specific position*. 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 such two-dimensional projections, the number of crossings does depend on the direction of the projection.

We note that strand fluctuations do not conserve the number of crossings; due to fluctuations, crossings disappear and appear and disappear over time. This appearance and disappearance will turn out to be related to virtual particles.

The fundamental principle declares that events are not points on manifolds; instead,

▷ ***Events*** are (one or several) 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. The strands are featureless: they have no mass, no tension, no stiffness, no branches, no fixed length, no ends, and they cannot be pulled, 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 are they? We will discuss this issue shortly.

Page 350

Page 165

▷ ***Strands*** are one-dimensional curves in three-dimensional space that reach the border of space.

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.

Imagining the strands as having Planck diameter does not make them observable, as this measurement result cannot be realized. (We recall that the Planck length is the lower bound on any length measurement.) In low energy situations, a vanishing strand diameter is an excellent approximation.

▷ In a purist definition, featureless strands have no diameter – neither the Planck length nor zero. They are better thought as *long thin clouds*.

Strands are unobservable and featureless, and thus have no diameter. Due to shape fluctuations, or equivalently, due to the shivering of space-time, the strands can be thought as having an effective diameter, akin to the diameter of a long thin cloud; this effective diameter is just a guide to our thinking. Since it is due to the shivering of the background space-time, the strand diameter is invariant under boosts. Funnels, mentioned below, might be a better visualization of the purist definition of strand. To keep this introduction as intuitive as possible, however, we stick with the idea of strands having an effective, invariant Planck diameter.

Page 169

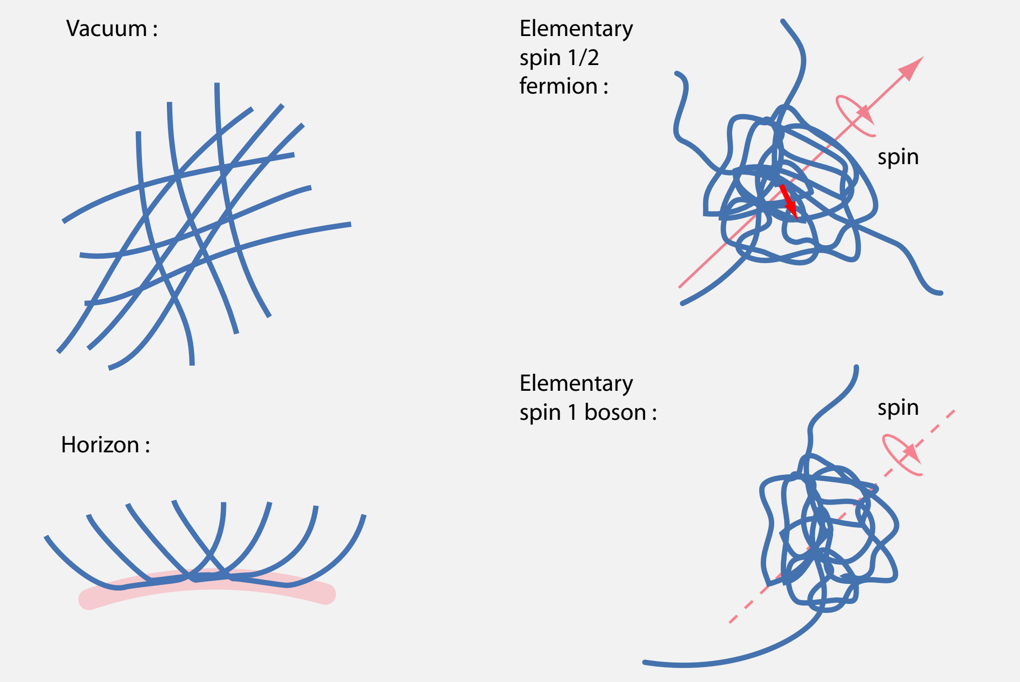


FIGURE 13 A first illustration of the basic physical systems found in nature; they will be explored in detail below.

The strand model distinguishes *physical space* from *background space*. We will show shortly why both concepts are required. With this distinction, the strand model asserts that matter and radiation, vacuum and horizons, are all built from *fluctuating strands* in a *continuous background*. We first clarify the two basic space concepts. Page 167

- ▷ ***Physical space***, or vacuum, is a physical system made of tangles that has size, curvature and other measurable properties.
- ▷ ***Continuous background space*** is introduced by the observer only to be able to describe observations. Every observer introduces his own background. It does not need to coincide with physical space, and it does not do so at the location of matter or black holes. But every observer's background is continuous and has 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; larger than any physical scale under discussion. In most situations of everyday life, when space is flat, background space and physical space coincide. However, they differ in situations with curvature and at Planck energy. Page 205

- ▷ ***Fluctuations*** change the position, shape and length of strands; fluctuations thus change position, orientation and phase of strand crossings. However,

TABLE 7 Correspondences between all known physical systems and mathematical tangles.

Physical system	Strand content	Tangle type
Vacuum and dark energy	many unknotted and untangled infinite strands	unlinked, trivial tangle
Graviton	two infinite twisted strands	rational tangle
Gravity wave	many infinite twisted strands	many rational tangles
Horizon	many woven infinite strands	woven, web-like tangle
Elementary vector boson (radiation)	one infinite curved strand	any one from a family of tangled curves
Classical electromagnetic wave (radiation)	many infinite curved strands	many helically deformed/tangled curves
Elementary quark (matter)	two infinite linked strands	rational tangle
Elementary lepton (matter)	three infinite linked strands	braided tangle

fluctuations never allow one strand to pass through another.

All strand fluctuations are possible, as long as strands do not interpenetrate. For example, there is no speed limit for strands. Whenever strand fluctuations lead to a crossing switch, they lead to an observable effect – be it a vacuum fluctuation, a particle reaction or a horizon fluctuation.

▷ Fluctuations are a consequence of the embedding of strands in a continuous background.

In the strand model, even isolated physical systems are surrounded by a bath of fluctuating vacuum strands. The properties of fluctuations, such as their spectrum, their density etc., are fixed once and for all by the embedding. Fluctuations are necessary for the self-consistency of the strand model.

Due to the impenetrability of strands – which itself is a consequence of the embedding in a continuous background – any disturbance of the vacuum strands at one location *propagates*. We will see below what disturbances exist and how they differ from fluctuations.

Fluctuating strands that lead to crossing switches 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 all possible strand configurations, in particular, all states that differ from flat vacuum states. By doing so, we will be able to classify every process and every system that we observe in nature.

We will discover that *all physical systems* can be constructed from strands. Table 7 gives a first overview of how vacuum, particles and horizons result from strand *tangles*.

▷ A ***tangle*** is a configuration of one or more strands that are *linked* or *knotted*. Tangles are characterized by their topology, i.e., by the precise way that they are linked or knotted.

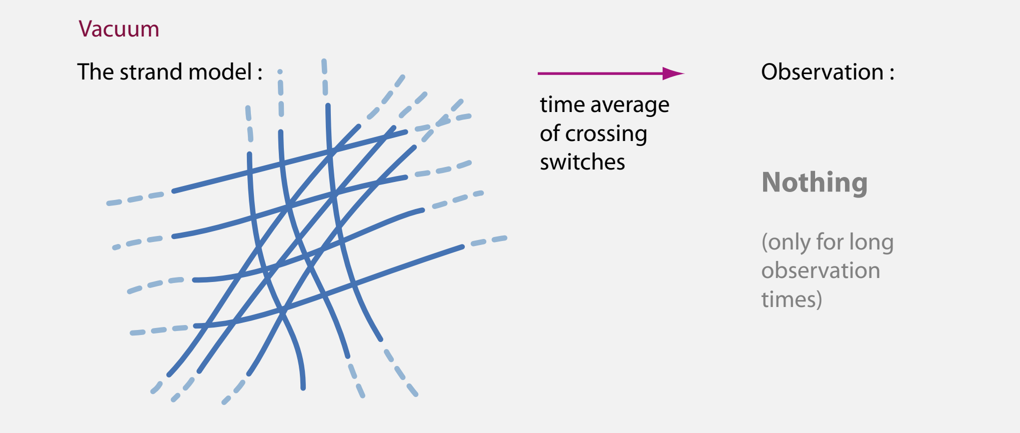


FIGURE 14 A illustration of the strand model for the vacuum.

Page 154 Some examples of important tangles are given in Figure 13. They will be discussed in detail in the following. Among others, we will discover that knots and knotted tangles do not play a role in the strand model.

We observe that vacuum, matter and radiation are all made of the *same* fundamental constituents, as required for a final theory. Page 311 We will discover below that classifying localized tangles naturally leads to the elementary particles that make up the standard model of particle physics – and to no other elementary particle.

We will also discover that strand fluctuations and the induced crossing switches in every physical system lead to the evolution equations and the Lagrangians of quantum field theory and of general relativity. In this way, strands describe *every* physical process observed in nature, including *all known interactions* and every type of motion.

Page 150 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 the one hand yields space-time, with its continuity, local isotropy and curvature, and on the other hand realizes the known connection between the quantum of action, spin and rotation.

VACUUM

We now construct, step by step, all important physical systems, concepts and processes from tangles. We start with the most important.

▷ ***Vacuum***, or ***physical space***, is formed by the time average of many unknotted fluctuating strands.

Figure 14 visualizes the definition. In the following, vacuum and physical space are always taken to be synonyms; the exploration will show that this is the most sensible use of the two concepts.* However, as mentioned, the strand model distinguishes *physical*

---

* We recall that since over a century, the concept of aether is superfluous, because it is indistinguishable from the concept of vacuum.

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 shortly, flat vacuum shows, averaged over time, no knots and no tangles, so that it is observed to be empty of matter and radiation. Temporary tangles that appear for a short time through vacuum fluctuations will be shown later to represent virtual particles. Page 235

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. The strand model thus replaces what used to be called 'space-time foam' or 'quantum foam'. Page 277

We also note that Figure 14 implies, despite appearances, that vacuum is *isotropic*. To see this, we need to recall that the observables are the crossing switches, not the strands, and that the observed vacuum isotropy results from the isotropy of the time-averaged strand fluctuations.

▷ We do not make any statement on the numerical density of strands in vacuum, or, equivalently, on their average spacing. Since strands are not observable, such a statement is not sensible. In particular, strands in vacuum are *not* tightly packed.

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 the alleged contradiction between continuity and minimum length are put to rest. In particular, physical space is *not* fundamentally discrete: a minimum length appears, though only in domains where physical space is undefined. On the other hand, the continuity of physical space results from an averaging process. Therefore, physical space is *not* fundamentally continuous: the strand model describes physical space as a homogeneous distribution of crossing switches. This is the strand version of Wheeler's idea space-time foam.

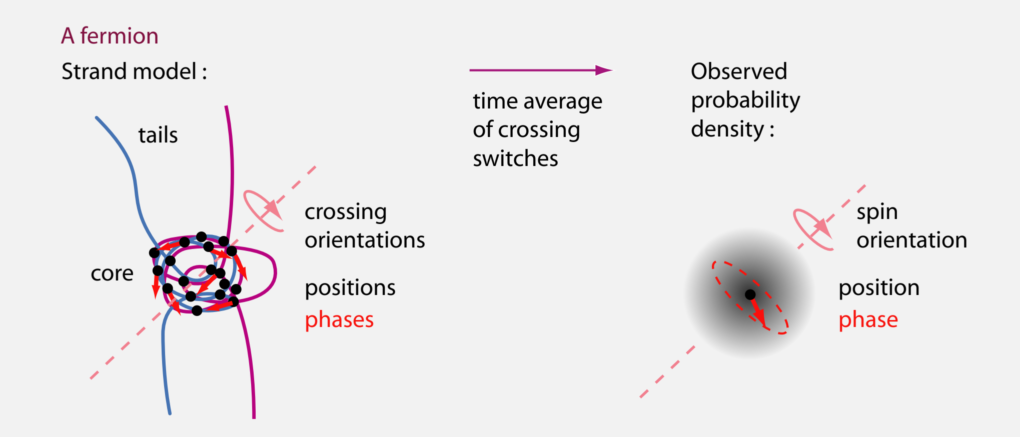
OBSERVABLE VALUES AND LIMITS

The fundamental principle implies the following definitions of the basic observables:

▷ The ***distance*** between two particles is the maximum number of crossing switches that can be measured 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 can be measured 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



Page 171 FIGURE 15 The tangle model of a spin 1/2 particle. More details will be given below.

state is the number of crossing switches that can be measured. 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 logarithm of the number of possible measurable crossing switches. 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, and thus continuous. Continuity is thus reconciled with the existence of a minimum measurable length and time interval.

Finally, we note that defining observables with the help of crossing switches automatically makes the Planck units c, $\hbar$, $c^4/4G$, k and all their combinations both *observer-invariant* and *limit* values. All these conclusions agree with the corresponding requirements for a final theory of nature. Page 147

PARTICLES AND FIELDS

Strands also define particles, as illustrated in Figure 15:

▷ 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 173

As shown in more detail soon, this definition of quantum particles yields, depending on the tangle details, either fermion or boson behaviour, and reproduces the spin–statistics theorem.

Page 221

Boson tangles will allow us to model field intensities. In particular, boson tangles allow us to deduce the electromagnetic and the two nuclear fields, as well as the corresponding gauge symmetries of the standard model of particle physics.

Page 211

Page 312

Modelling fermions as tangles will allow us to deduce Dirac's equation for relativistic quantum particles (and the Schrödinger equation for non-relativistic particles). Still later, by classifying all possible tangles, we will discover that only a *finite* number of possible elementary particles exist, and that the *topological type* of tangle determines the mass, mixings, quantum numbers, charges and couplings of each elementary particle. We could also speak of a *tangle model*.

In the 1960s, John Wheeler stated that a unified description of nature must explain 'mass without mass, charge without charge, field without field'. The strand model realizes this aim, as we will find out.

Before we deduce modern physics, we first take a break and explore some general issues of the strand model.

CURIOSITIES AND FUN CHALLENGES ABOUT STRANDS

Page 225

Why do crossing switches have such a central role in the strand model? An intuitive explanation follows from their role 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, in the end, detection of photons. And the strand model shows that photon absorption and detection are intimately related to the crossing switch, as we will find out below.

* *

Is there a limit to the fluctuations of strands? Yes and no. On the one hand, the 'speed' of fluctuations is unlimited. On the other hand, fluctuations with a 'curvature 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.

* *

Challenge 118 e

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 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 302 in the strand model, the universe is made of a single strand folded in a complicated way. Nature is one strand. Therefore, strands are not made of something, they are made of everything. The substance of strands is nature itself.

* *

Since there is only one strand in nature, strands are not a reductionist approach. At Planck scale, nature is one and indivisible.

* *

What are particles? In the strand model, elementary particles are (families of) *tangles* of strands. In other words, elementary particles are not the basic building blocks of matter – strands are. If particles could *really* be elementary, it would be impossible to understand their properties.

In the strand model, particles are not really elementary, but neither are they, in the usual sense, composed. Particles are tangles of unobservable strands. In this way, the strand model retains the useful aspects of the idea of elementary particle but gets rid of its limitations. In a sense, the strand model can be seen as eliminating the concepts of elementariness and of particle. This confirms and realizes another requirement that we had deduced earlier on. Page 77

* *

Challenge 119 e Can macroscopic determinism arise at all from randomly fluctuating strands?

* *

Challenge 120 s Do parallel strands form a crossing? Do two distant strands form a crossing?

* *

Challenge 121 s Is a crossing switch defined in more than three dimensions?

* *

Challenge 122 s Can you find a way to generalize or to modify the strand model?

* *

Looking back, we might note a relation between the strand model and the vision of ‘it from qubit’ that is propagated by David Deutsch. What is the difference between the fundamental principle and a qubit? Challenge 123 e

* *

Is the strand model confirmed by other, independent research? Yes, a few years after the strand model appeared, this started to happen. For example, 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 result that ‘space at a fixed time is thus threaded by rapidly fluctuating lines’. This is exactly what the strand model states. Ref. 154

Other theoretical approaches that confirm the strand model are mentioned in various places later in the text. Despite such developments, the essential point remains to check Page 297 how the strand model compares with experiment. Given that the strand model turns out to be unmodifiable, there are no ways to amend predictions that turn out to be wrong. If a single prediction of the strand model turns out to be incorrect, the model is doomed. But so far, no experimental prediction of the strand model contradicts experiments.

Page 393

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 matter at Planck scales: both are made of strands. The strand model implies that observables are not real numbers at Planck scales. The strand model implies that the universe 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 not clearly defined and that nature is not a set. You can check by yourself that all other paradoxes appear automatically. Furthermore, almost all requirements for a final theory listed in Table 6 are fulfilled. Only two requirements of the table must be discussed in more detail: the requirements of complete precision and of unmodifiability. We start with complete precision.

Page 64
Page 73
Page 90
Page 127
Challenge 124 e

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. This requires, first of all, that the model describes *all* experiments. As will be shown below, this is indeed the case, because the strand model contains both general relativity and the standard model of particle physics. But secondly and most importantly, the model must also settle all those questions that were left unanswered by twentieth-century fundamental physics. Because the questions, the *millennium list* of open issues, are so important, they are given, again, in Table 8.

TABLE 8 The millennium list: *everything* the standard model 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 since the year 2000
Local quantities unexplained by the standard model: particle properties	
$\alpha = 1/137.036(1)$	the low energy value of the electromagnetic coupling or fine structure 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

TABLE 8 (Continued) The millennium list: *everything* the standard model and general relativity *cannot* explain; also the *only* experimental data available to test the final, unified description of motion.

Observable	Property unexplained since the year 2000
$3 \cdot 4$	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
Concepts unexplained by the standard model	
$c, \hbar, k$	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
Ψ	the origin and nature of wave functions
$S(n)$	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, of the vanishing of magnetic charge, and of minimal coupling
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
Renorm. group	the origin of renormalization properties
$\delta W = 0$	the origin of 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 unexplained by general relativity and cosmology	
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)
$\rho_{\mathrm{de}} = \Lambda c^4/(8\pi G)$ $\approx 0.5\,\mathrm{nJ/m^3}$	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
ρ_{dm}	the density and nature of dark matter
$f_0(1, ..., c.\ 10^{90})$	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
Concepts unexplained by general relativity and cosmology	
c, G	the origin of the invariant Planck units of general relativity
$\mathrm{R} \times \mathrm{S}^3$	the observed topology of the universe
$G^{\mu\nu}$	the origin and nature of curvature, the metric and horizons
$\delta W = 0$	the origin of the least action principle in general relativity
$W = \int L_{\mathrm{GR}} \mathrm{d}t$	the origin of the Lagrangian of general relativity

The open issues in the millennium list must be resolved by any final, unified model of nature, and thus also by the strand model. The open issues can be summarized in two general points:

— *Reproduce* quantum theory, the standard model, general relativity and cosmology.

— *Explain* masses, mixing angles and coupling constants.

Of course, only the second point is the *definite test* for a final, unified description. But we need the first point as well. The following chapters deal with both points.

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. To be a full explanation, such a description must not only be logically and experimentally complete, it must also be *unmodifiable*. Even though Einstein made the point for general relativity, this important aspect is rarely discussed with clarity. In particular, any unmodifiable explanation has two main properties: first, it cannot be generalized, and second, it is not itself a generalization.

Generalizing models is a sport among theoretical and mathematical physicists. 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 complicated gauge symmetries, with different types of supersymmetry, with more Higgs bosons, or with additional heavy neutrinos. In the case of the strand model, researchers will also explore models with more complicated entities than strands, such as bands or bifurcating entities, and any other generalization they can imagine.

▷ Can a *final* description of nature have generalizations? The answer is *no*.

Indeed, if it were possible to generalize the final 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 theoretical and mathematical physicists just described must all be impossible. So far, investigations confirm this prediction: no generalization of the strand model has been found yet.

Where does this fondness for generalization come from? In the 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, generalizing 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. This connection leads to a second point.

General relativity and the standard model of particle physics must indeed be *approx-*

imations of the final theory. But can either general relativity or the standard model be *special cases* of the final, unified theory? Or, equivalently:

▷ Can the unified theory itself be a generalization of existing theories? The answer is again *no*.

Because neither general relativity nor the standard model of particle physics are able to explain the millennium issues, any generalization of them would also be unable to do so. Generalizations have *no* explanatory power. If the unified theory were a generalization of the two existing theories, it could not explain any of the millennium issues of Table 8! Therefore, general relativity and the standard model of particle physics must be approximations, but *not* special cases, of the final theory. In particular, if the strand model is a final description, *approximations* of the strand model must exist, but *special cases* must not. This is indeed the case, as we will find out.

To summarize, a final theory must be an *explanation* of all observations. An explanation of an observation is the recognition that it follows unambiguously, *without alternative*, from a general property of nature. We conclude that the final, unified description of motion must neither allow generalization nor must it be a generalization of either the standard model or 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 eliminates grand unification, supersymmetry and higher dimensions as aspects of the final theory: indeed, these ideas *generalize* the standard model of elementary particles and they are *modifiable*. Therefore, all these ideas lack explanatory power.

Challenge 125 e

A final and unified theory must be an *unmodifiable explanation* of general relativity and the standard model. Because neither supersymmetry, nor the superstring conjecture, nor loop quantum gravity explain the standard model of particle physics, they are not unified theories. Because these models are modifiable, they are not final theories. In fact, at least one of these two aspects is lacking in every candidate final theory proposed in the twentieth century.

We will discover below that *the strand model is unmodifiable*. Its fundamental principle cannot be varied in any way without destroying the whole description. Indeed, no modification of the strand model or of the fundamental principle has been found so far. We will also discover that the strand model *explains* the standard model of particle physics and explains general relativity. The strand model is thus a candidate for the final theory.

Challenge 126 e

Ref. 155

* Independently, David Deutsch made 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.

WHY STRANDS? – SIMPLICITY

> Simplex sigillum veri.*
>
> Antiquity

Let us assume that we do not know yet whether the strand model can be modified or not. Two other reasons still induce us 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, sets or any axiomatic system. Let us explore the issue of simplicity first. Page 105

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, because 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 is impossible in three spatial dimensions. Page 115

Ref. 140 No *simpler* model than featureless strands is possible. All other extended constituents that have been explored – ribbons, bands, strings, membranes, posets, branched lines, networks, crystals and quantum knots – increase the complexity of the model. In fact these constituents increase the complexity in *two* ways: they increase the number of features of the fundamental constituents and they complicate the mapping from the model to observation. Ref. 156 Ref. 157 Ref. 158

First, no other model based on extension uses *featureless* constituents. In all other models, the fundamental constituents have properties such as tension, field values, coordinates, quantum numbers, shape, twists, orientation, 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 allow alternatives and are difficult if not impossible to justify. In addition, these features increase the complexity of the possible processes. In contrast, the strand model has no justification issue and no complexity issue.

Secondly, the link between more complicated models and experiment is often *intricate* and sometimes not unique. As an example, the difficulties to relate superstrings to experiments 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 structure of nature at Planck scales. The comparable ideas in other models are much more elaborate. Ref. 150

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*,

---

[*] 'Simplicity is the seal of truth.'

TABLE 9 The differences between nature and any description.

Nature	Description
Nature is not a set.	Descriptions need 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 but contains circular definitions.	Axiomatic descriptions are needed for precise talking and thinking.

we need concepts such as sets, neighbourhoods, 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 fact, we will discover that strands indeed are the *simplest* way to model particles, interactions and the vacuum, while fulfilling the requirements of a final theory.

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 can be checked against observation. In particular, a simple model allows simple checking of its solution of paradoxes. Above all, we can resolve the most important paradox of physics.

WHY STRANDS? – THE FUNDAMENTAL CIRCULARITY OF PHYSICS

> “Without the concepts *place*, *void* and *time*, change cannot be. [...] It is therefore clear [...] that their investigation has to be carried out, by studying each of them separately.”
>
> Aristotle *Physics*, Book III, part 1.

The strand model describes strands as fluctuating in a background space-time of three plus one space-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 108

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 the one hand, as talkers, we must use *concepts that allow us to talk*. Talking and thinking requires that we use continuous space and time: on short, we must use a *background*. The background must be continuous, without minimum length. 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 listed in Table 9. We know that nature is not a set, has no points, no point particles and no locality, but that it is dual. But in order to talk about nature, we need a background that lacks all these properties. Because there is no way to get rid of these apparently contradictory requirements, we don't:

▷ We use both continuous background space-time and discrete strands to describe nature.

In a few words: A unified model of physics allows talking about motion with highest precision; this requirement forces us to use, *at the same time*, both continuous space-time and discrete strands. This double use is not a contradiction but, as just explained, the result of a *circular definition*. Since we, the talkers, are part of nature, a unified model means that we, the talkers, talk about ourselves.

We stress that despite the circularity of physics, Gödel's incompleteness theorem does not apply to the situation. In fact, the theorem does not apply to any unified theory of physics for two reasons. First, the incompleteness theorem applies to *self-referential* statements, not to circular definitions. Vol. III, page 303 Self-referential statements do not appear in physics, not in sensible mathematics and not in the strand model. Secondly, Gödel's theorem applies to mathematical structures based on sets, and the final theory is not based on sets.

We do not state that *background* space and time exist *a priori*, as Immanuel Kant states, Ref. 159 but only that background space and time are *necessary* for thinking and talking, as Aristotle states. In fact, *physical* 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. Figure 16 illustrates the solution proposed by the strand model.

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

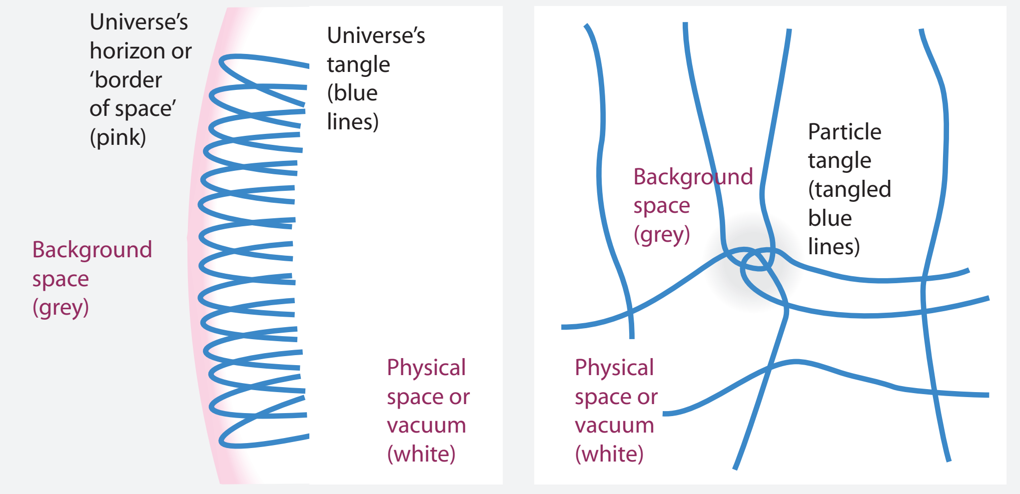


FIGURE 16 In the strand model, physical space – or vacuum – and background space are distinct, both near the horizon and near particles.

we attempt to define *sets* or *elements* (or points) 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 – i.e., for the description of all motion – and there cannot be one. Page 108

A further issue must be discussed in this context. As mentioned, strands fluctuate in a background space, and only crossing switches can be observed. In particular, this implies 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, 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 a further 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 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 Ref. 160

FIGURE 17 Two equivalent depictions of the fundamental constituents of nature: strands and funnels.

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.* Many 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, and only allows points or sets as approximate concepts.

FUNNELS – AN EQUIVALENT ALTERNATIVE TO STRANDS

Another type of constituent also fulfils all the conditions for a unified description. As shown in Figure 17, as an alternative to fluctuating strands, we can use fluctuating *funnels* as fundamental constituents. In the description with funnels, nature resembles a complicated tangle of a three-dimensional space that is projected back into three dimensions.

Funnels show that the strand model only requires that the effective *minimal* effective diameter of a strand is the Planck length; it could have other diameters as well. Funnels also show that due to varying diameters, strands can, through their fluctuations, literally be everywhere in space and thus effectively *fill* space, even if their actual density is low.

Funnels resemble many other research topics. Funnels are similar to wormholes; however, both their ends lead, at the border of space, 'into' usual three-dimensional space. Funnels are also similar to D-branes, except that they are embedded in three spatial dimensions, not ten. Funnels also resemble a part of an exotic manifold projected into three-dimensional space. Fluctuating funnels also remind us of the amoeba mentioned above. However, the similarities with wormholes, D-branes or exotic manifolds are of little help: so far, none of these approaches has led to viable models of unification.

Page 129

A first check shows that the funnel alternative seems *completely equivalent* to strands.* You might enjoy testing that all the conclusions deduced in the following pages appear

---

* Two issues that put this equivalence into question are ending funnels and diameter behaviour under boosts. The first issue is subject of research, but it is expected that it poses no problem. The second issue is mitigated by the shivering of the background space.

Challenge 127 e

unchanged if strands are replaced by funnels. In particular, also funnels allow us to deduce quantum field theory, the standard model and general relativity. 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.

Knots and the ends of strands

In the original strand model, developed in the year 2008, strands that contain knots were part of the allowed configurations. This has the disadvantage that the formation of a knot requires at least one loose end that is pulled through a strand configuration. Such loose ends, however, go against the aim of highest possible simplicity of the unified model.

Later, in 2015, it became clear that strands without knots are sufficient to recover modern physics. This is shown in the following. For this reason, knots and the ends of strands play no role in the model any more; the aim of highest possible simplicity is now realized.

SUMMARY ON THE FUNDAMENTAL PRINCIPLE – AND ON CONTINUITY

We have introduced featureless, fluctuating *strands* as common constituents of space, matter, radiation and horizons. We defined fundamental events as crossing switches of strands. All physical processes are composed of fundamental events. Events and the values of all physical *observables* are defined with the help of Planck units, which in turn are due to crossing switches of strands. The definition of all physical observables through Planck units with the help of crossing switches of strands is the *fundamental principle.*

Using the fundamental principle, *continuity* of any kind – of space, fields, wave functions or time – results from the time averaging of crossing switches. This issue will be explored in detail below.

Page 147

The strand model fulfils the general requirements for the final and unified description listed in Table 6, *provided* that it describes all motion with full precision and that it is unmodifiable.

Page 161

At this point, therefore, we must start the comparison with experiment. We need to check whether strands describe *all* motion with *complete* precision. Fortunately, the task is limited: we only need to check whether strands solve each of the millennium issues listed in Table 8. If the strand model can solve those issues, then it reproduces all observations about motion and provides a final and unified description of nature. If the issues are not solved, the strand model is worthless.

CHAPTER 8

QUANTUM THEORY OF MATTER DEDUCED FROM STRANDS

We show in this chapter that featureless strands that fluctuate, together with the fundamental principle – defining $\hbar/2$ as a crossing switch – imply without alternative that matter is described by quantum theory. More precisely, we deduce that tangles of fluctuating strands reproduce the spin 1/2 behaviour of matter particles, allow us to define wave functions, and imply the Dirac equation for the motion of matter. In particular, we first show that the components and phases of the wave function at a point in space are due to the orientation and phase of strand crossings at that point. Then we show that the Dirac equation follows from the belt trick (or string trick).

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. So far, it seems that the strand model is the only microscopic model of relativistic quantum theory that is available in the research literature.

In the present chapter, we derive the quantum theory of *matter*: we show that strands reproduce all observations about fermions and their motion. We leave for later the derivation of the quantum theory of light and the nuclear interactions, the standard model of elementary particles, and the quantum description of gravitation. As usual in quantum theory, we work in *flat* space-time.

STRANDS, VACUUM AND PARTICLES

In nature, particles move in the vacuum. The vacuum is free of matter and energy. In the strand model,

▷ ***Vacuum*** is a collection of fluctuating, unknotted and untangled strands.

Page 156
Page 235

The vacuum is illustrated in Figure 14. The time average of unknotted and untangled strands has no energy and no matter content, because there are – averaged over time – *no* crossing switches and *no* tangles. The temporary crossing switches that can appear through fluctuations of the vacuum will turn out to be virtual particles; we will explore them below. We note that the physical vacuum, being a time average, is *continuous*. The flat physical vacuum is also *unique*: it is the same for all observers. The strand model thus contains both a minimum length and a continuous vacuum. The two aspects do not contradict each other.

In nature, quantum particles *move*: quantum particles change position and phase over

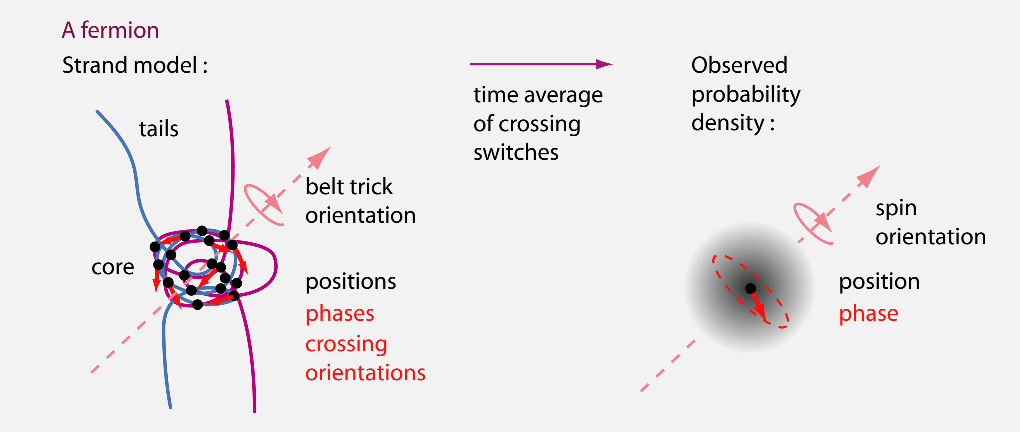


FIGURE 18 A fermion is described by a tangle of two or three strands. The crossings in the tangle core and their properties lead, after averaging, to the wave function and the probability density.

time. We therefore must define these concepts. At this stage, as just explained, we concentrate on quantum matter particles and leave radiation particles for later on. As illustrated in Figure 18 and Figure 19, we define:

- ▷ An elementary ***matter particle***, or ***fermion***, is a tangle of two or more strands that realizes the belt trick.

The details of this definition will become clear shortly, including the importance of the related tangle *family*. In every tangle, the important structure is the *tangle core*: it is the tangled or knotted part of the tangle that contains all the links and crossings. The core is connected to the border of space by the *tails* of the tangle.

- ▷ The ***position*** of a particle is given by the centre of the averaged tangle core. The particle position is thus the average of all its crossing positions.
- ▷ The ***phase*** of a matter particle is given by half the angle that describes the orientation of the tangle core around the spin axis. The particle phase is thus the average of all its crossing phases.
- ▷ The ***spin orientation*** of a matter particle is given by the rotation axis of the core. The spin orientation is thus the average of all its crossing orientations.
- ▷ The ***wave function*** of a matter particle is a blurred rendering of the crossing of its fluctuating strands.

These definitions are illustrated in Figure 18 and will be explored in detail below. We note that all these definitions imply a short-time average over tangle fluctuations. With the definitions, we get: Page 183

- ▷ ***Motion*** of any quantum 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 the appearance of the invariant quantum of action $\hbar$. 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 in detail *how* the quantum of action $\hbar$ determines the motion of quantum particles. In particular, we will show that tangle fluctuations reproduce usual textbook quantum theory. As an advance summary, we clarify that

▷ ***Free quantum particle motion*** is due to fluctuations of tangle *tails*. The deformations of the tangle core are not important for free motion, and we can neglect them in this case.

In other words, when exploring quantum theory, we approximate tangle cores as being *rigid*. We will study *core deformations* in the next chapter, where we show that they are related to *interactions*. Core deformations will lead to quantum *field* theory. In this chapter we explore just the deformations of tangle *tails*; they produce the motion of *free* (and stable) quantum particles. In short, tail deformations lead to quantum mechanics. To deduce quantum mechanics from strands, we first study the rotation and then the translation of free matter particles.

Page 221

ROTATION, SPIN 1/2 AND THE BELT TRICK

In nature, quantum particles are described by their behaviour under *rotation* and by their behaviour under *exchange*. The behaviour of a particle under rotation is described by its spin value, its spin axis and its phase. The behaviour of quantum particles 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*.

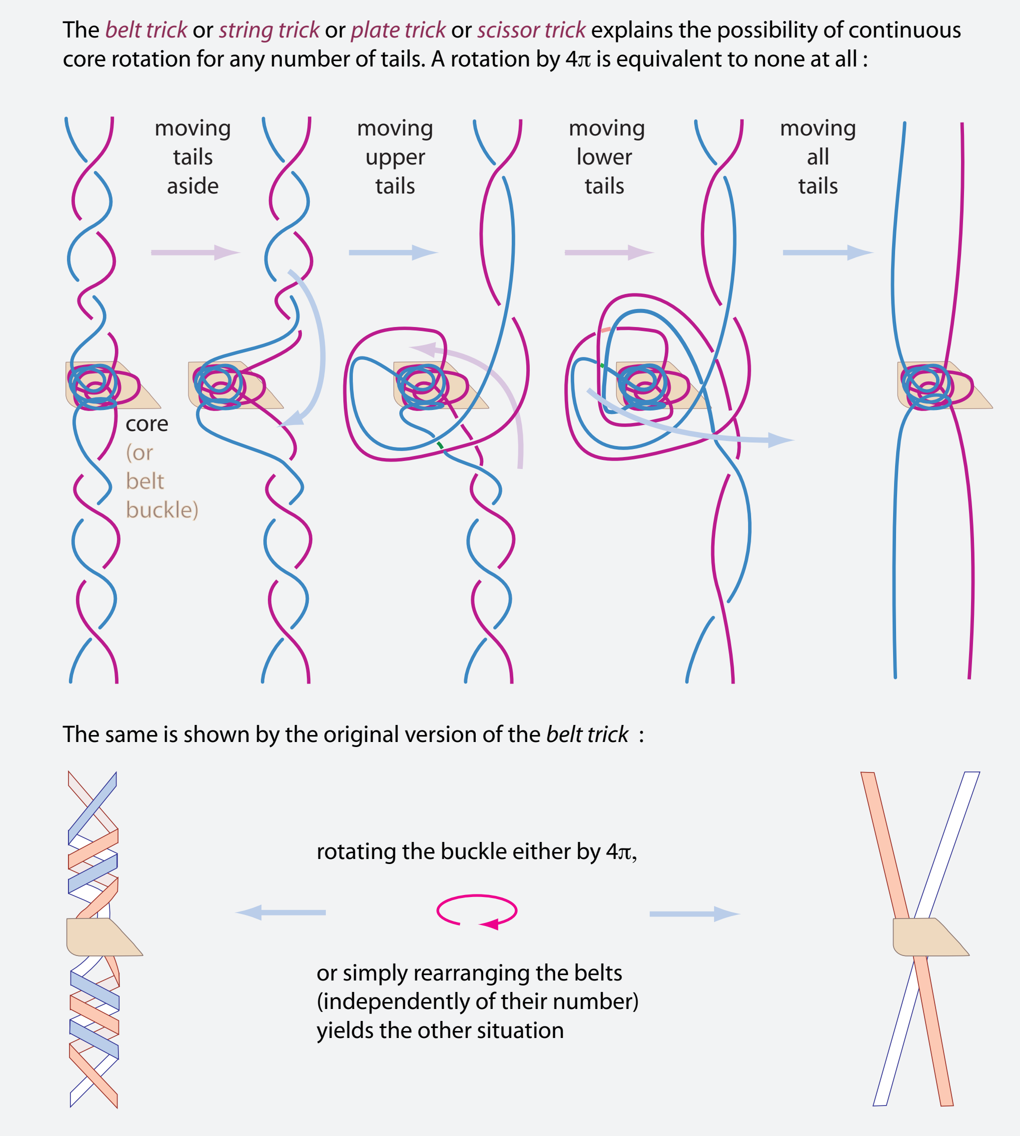
We now show that all properties of particle rotation and exchange follow from the strand model. We start with the case of spin 1/2 particles, and first clarify the nature of particle rotation. (We follow the usual convention to use ‘spin 1/2’ as a shorthand for ‘z-component of spin with value $\hbar/2$’.)

It is sometimes claimed 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, it is said, spin 1/2 particles cannot be modelled as small rotating stones. Secondly, it is allegedly 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. These statements are correct. Despite being correct, there is a way to get around them, namely by modelling particles with strands; at the present stage, we focus on the first argument: 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:


The *belt trick* or *string trick* or *plate trick* or *scissor trick* explains the possibility of continuous core rotation for any number of tails. A rotation by 4π is equivalent to none at all :

The same is shown by the original version of the *belt trick* :

rotating the buckle either by 4π,

or simply rearranging the belts (independently of their number) yields the other situation

FIGURE 19 The *belt trick* – or *string trick* or *plate trick* or *scissor trick* – shows that a rotation by 4 π of a central object with three or more tails (or with one or more ribbons) attached to spatial infinity is equivalent to no rotation at all. This equivalence allows a suspended object, such as a belt buckle or a tangle core, to rotate for ever. The belt trick thus shows that tangle cores made from two or more strands behave as spin 1/2 particles.

▷ ***Quantum particles*** are particles whose tails cannot be neglected.

For stones and other everyday objects, tails do not play an important role, because everyday objects are mixed states, and not eigenstates of angular momentum. In short, in everyday objects, tails can be neglected. Therefore, everyday objects are neither fermions nor bosons. But for quantum particles, the tails are essential. Step by step we will see

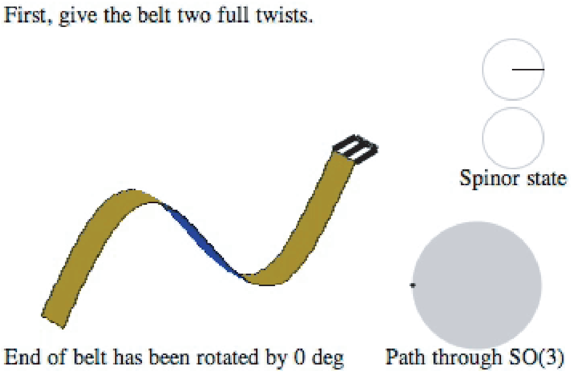


FIGURE 20 The belt trick: a double rotation of the belt buckle is equivalent to no rotation; te animation shows one way in which the belt trick can be performed. Not shown: the belt trick is also possible with *any* number of belts attached to the buckle. (QuickTime film © Greg Egan)

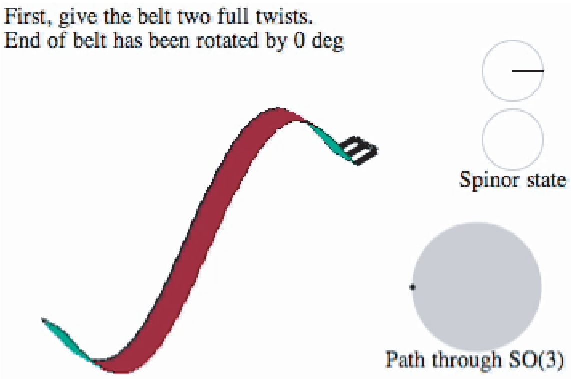


FIGURE 21 The belt trick again: this animation shows another way – another direction – in which the trick can be performed. Not shown: the belt trick is also possible with *any* number of belts attached to the buckle. (QuickTime film © Greg Egan)

that

▷ The tails of quantum particles explain their *spin behaviour*, their *exchange behaviour* and their *wave behaviour*.

In particular, we will see 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.

The spin behaviour of quantum particles is a consequence of strand tails. Indeed, it has been known for about a century that the so-called *belt trick* – illustrated in Figure 19, 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. 161

The belt trick is of central importance in the strand model of spin 1/2 particles. In

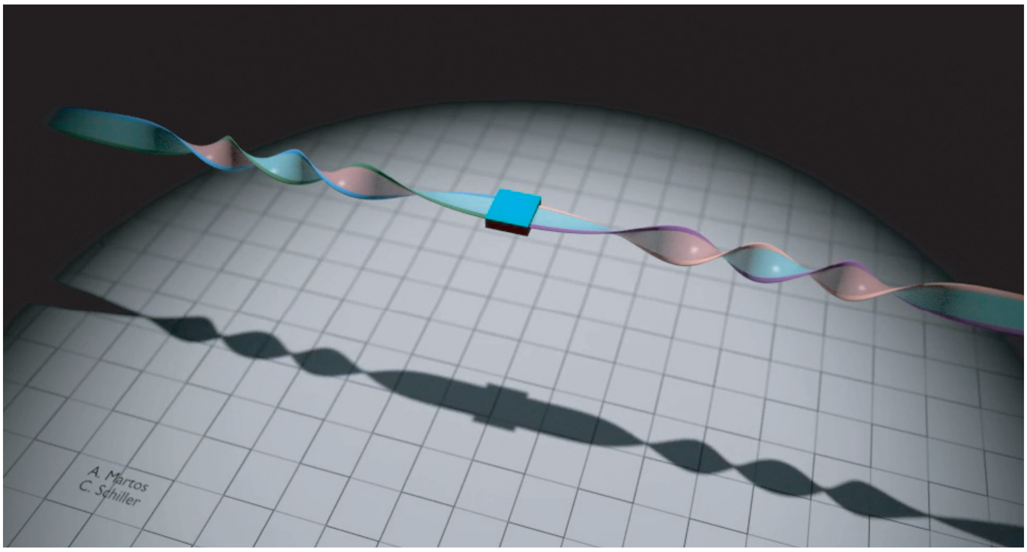


FIGURE 22 Assume that the belt cannot be observed, but the square object can, and that it represents a particle. The animation then shows that such a particle (the square object) can return to the starting position after rotation by 4π (but not after 2π). Such a 'belted' or 'tethered' particle thus fulfils the defining property of a *spin 1/2 particle*: rotating it by 4π is equivalent to no rotation at all. The belted square thus represents the spinor wave function; for example, a 2π rotation leads to a twist; this means a change of the sign of the wave function. A 4π rotation has no influence on the wave function. The equivalence is shown here with two attached belts, but the trick works with any positive number of belts! You can easily repeat the trick at home, with a paper strip or one or several real belts. (QuickTime film © Antonio Martos)

Page 197

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 19. 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 simple strands and thus indeed unobservable. We will show below that the intermediate twisting of the tails that appears after rotation by only 2π corresponds to a multiplication of the wave function by -1, again as expected from a spin 1/2 particle.

Page 188

If we replace each belt by its two coloured edges, Figure 22 shows how tails behave when a spin 1/2 tangle is rotated. By the way, systems with tails – be they strands or bands – are the *only possible systems* that realize the spin 1/2 property. Only systems with tails to spatial infinity have the property that a rotation by 4π is equivalent to no rotation at all. (Can you show this?) The fundamental connection between spin 1/2 and extension is one of the properties that led to the strand model.

Challenge 128 e

The animations show that the belt trick works with one and with two belts attached to the buckle. In fact, belt trick works with *any number* of belts attached to the buckle. The belt trick even works with *infinitely* many belts, and also with a *full two-dimensional sheet*. The wonderful video www.youtube.com/watch?v=UtdljdoFAwg by Gareth Taylor

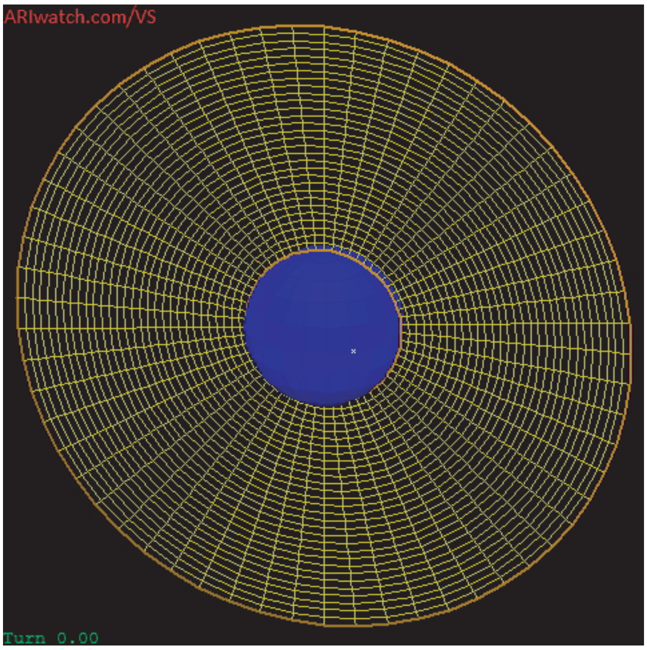


FIGURE 23 The belt trick realized as a rotating ball attached to a sheet (QuickTime film © www.ariwatch.com/VS/Algorithms/DiracStringTrick.htm).

and the slightly different animation of Figure 23 both illustrate the situation. A sphere glued to a flexible sheet can be rotated as often as you want: if you do this correctly, there is *no* tangling and you can go on for ever.

The animations of the belt trick lead us to a statement on strands and tangles that is central for the strand model:

▷ An object or a tangle core that is attached by (three or more) tails to the border of space can rotate *continuously*.

Here we made the step from belts to strands. In other terms, the possibility of continuous rotation allows us to describe spin 1/2 particles by rotating tangles. In other terms,

▷ Rotating tangles model spin.

The tail fluctuations required to rearrange the tails after two full turns of the core 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.

THE BELT TRICK IS NOT UNIQUE

Ref. 161 One aspect of the belt trick seems unmentioned in the research literature: after a rotation of the belt buckle or tangle core by 4π, there are various options to untangle the tails. Two different options are shown in Figure 20 and Figure 21. You can test this yourself, using a real belt. In fact, there are *two* extreme ways to perform the belt trick, and a continuum Challenge 129 e

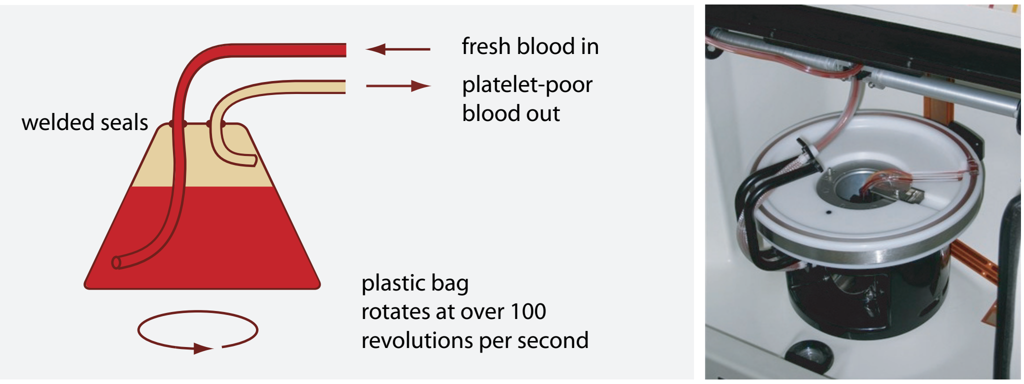


FIGURE 24 In an apheresis machine, the central bag spins at high speed despite being connected with tubes to a patient; this happens with a mechanism that continuously realizes the belt trick (photo © Wikimedia).

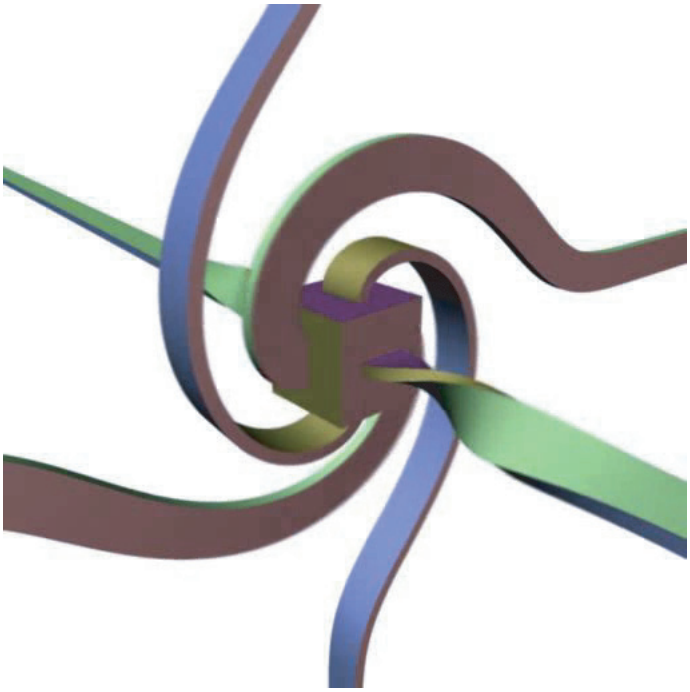


FIGURE 25 The basis of the apheresis machine – and yet another visualisation of the belt trick, here with 6 belts (QuickTime film © Jason Hise).

of options in between. These options will be of central importance later on: the options require a description of fermions with *four* complex functions. We will discover that the various options of the belt trick are related to the difference between matter and antimatter and to the parity violation of the weak interaction.

FIGURE 26 A version of the antitwister mechanism, or belt trick, with 96 belts attached to a black and white ball that rotates continuously (QuickTime film © Jason Hise).

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 then flows back into the other arm of the donor. This happens continuously, for about an hour or two. In order to be *sterile*, tubes and bag are used only once and are effectively *one* piece, as shown in Figure 24. Apheresis machines need *tethered rotation* to work. Topologically, this set-up is identical to a fermion tangle: each tube corresponds to one belt, or two strand tails, and the rotating bag corresponds to the rotating core.

In such apheresis machines, the centrifugation of the central bag takes place at over 100 revolutions per second, in the way illustrated in Figure 25. To avoid tangling up the blood tubes, a bracket moves the tubes during each rotation, alternatively up and down. This so-called *anti-twister mechanism* produces precisely the motion along which the belt moves when it is untangled after the buckle is rotated by 4π. An apheresis machine thus performs the belt trick 50 times per second, with each rotation of the centrifugation. Due to the centrifugation, the lighter platelets are retained in the bag, and the heavier components of the blood are pumped back to the donor. The retained platelets 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, without the belt trick, platelet donations would not be sterile and would thus be impossible. Only the belt trick, or tethered rotation, allows sterile platelet donations that save other people's lives.

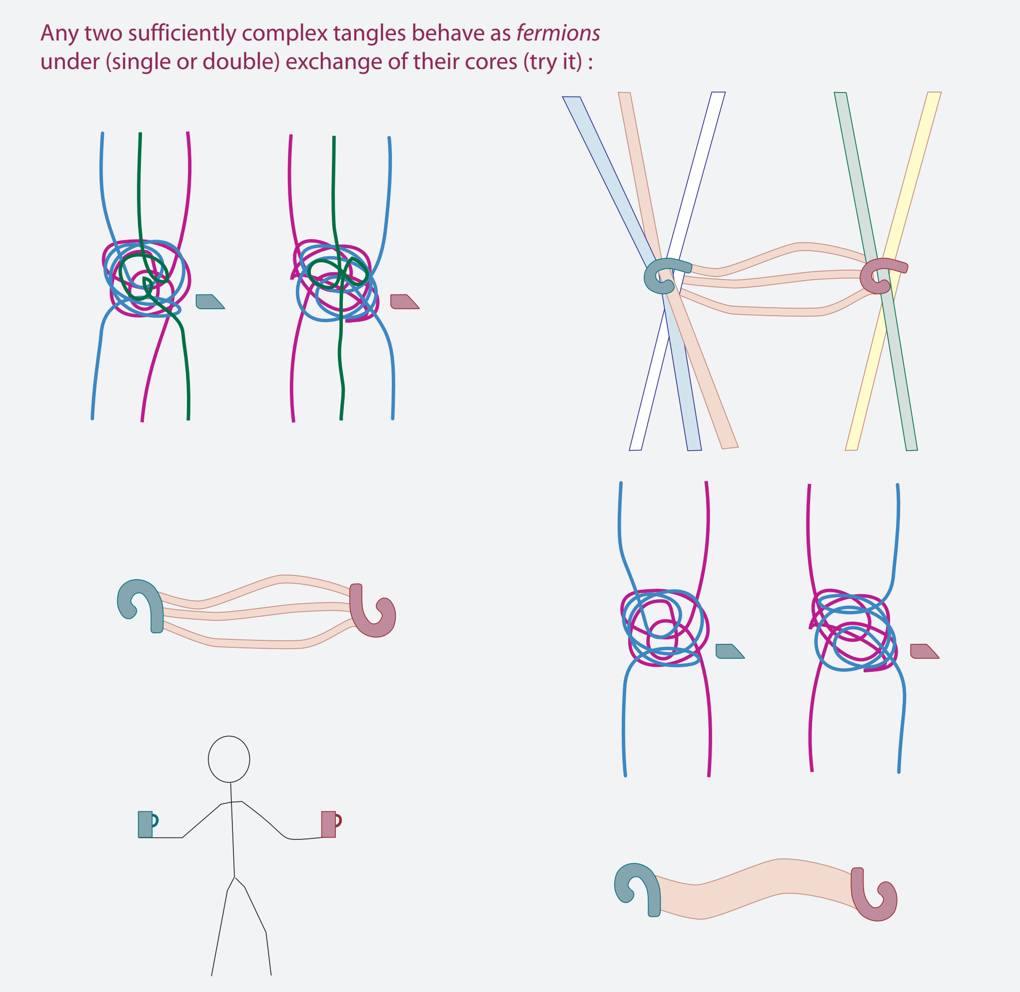


FIGURE 27 When two spin 1/2 tangles each made of several strands or bands, are exchanged *twice*, it is possible to rearrange their tails to yield the original situation. This is not possible when the tangles are only rearranged *once*. Spin 1/2 tangles are thus fermions. The figure presents the most common systems that show this behaviour.

Fermions and spin

In nature, *fermions* are defined as those particles whose wave function changes sign when they are exchanged. Does the strand model reproduce this observation?

We will see below that in the strand model, wave functions are *blurred* tangles. We thus can explore exchange properties of quantum particles and of their wave functions by exploring the exchange properties of their tangles. Now, 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! 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 27, or by watching the animation of Figure 28.

The simplest possible version of the experiment is the following: take two coffee cups,

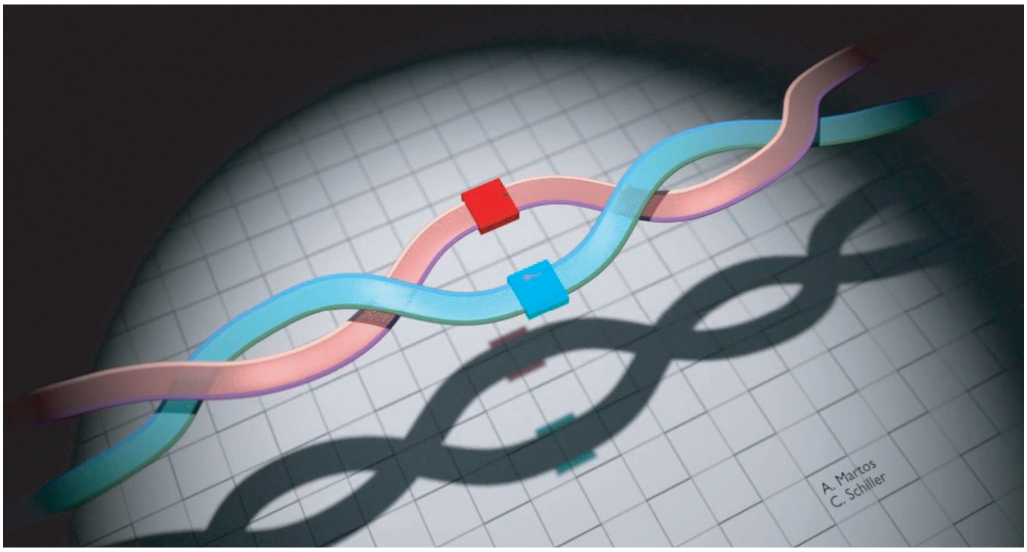


FIGURE 28 Assume that the belts cannot be observed, but the square objects can, and that they represent particles. We know from above that belted buckles behave as spin 1/2 particles. The animation shows that two such particles return to the original situation if they are switched in position twice (but not once). Such particles thus fulfil the defining property of *fermions*. (For the opposite case, that of bosons, a simple exchange would lead to the identical situation.) You can repeat the trick at home using paper strips. The equivalence is shown here with two belts per particle, but the trick works with any positive number of belts attached to each buckle. This animation is the essential part of the proof that spin 1/2 particles are fermions. This is called the *spin–statistics theorem*. (QuickTime film © Antonio Martos)

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: in this case, all arms return to the starting situation.

All these experiments show:

▷ A *simple* exchange of two spin 1/2 particles (tangles, cups on hands, belt buckles) is equivalent to a multiplication by −1, i.e., to *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 other words, the strand model reproduces the spin–statistics theorem for spin 1/2: all elementary matter particles are 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.

We note that it is sometimes claimed that the appearance of spin 1/2 can only be mod-

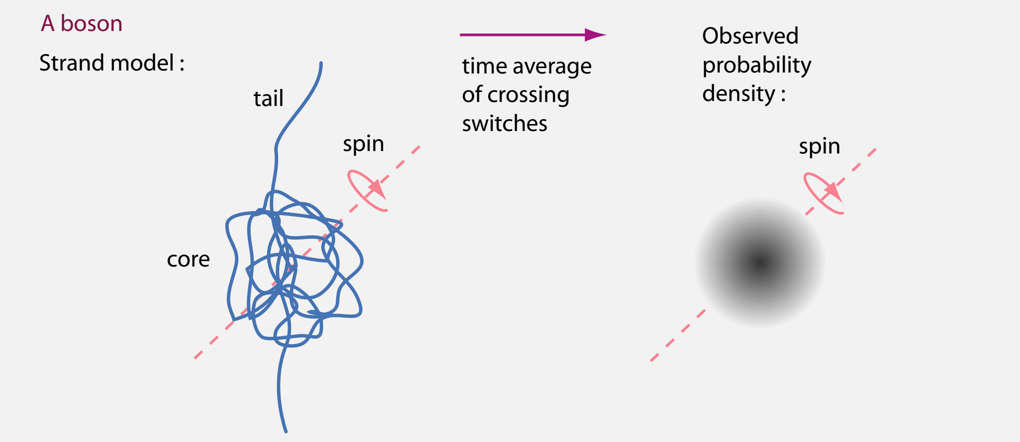


FIGURE 29 A massive spin 1 particle in the strand model (left) and the observed probability density when averaging its crossings over long time scales (right).

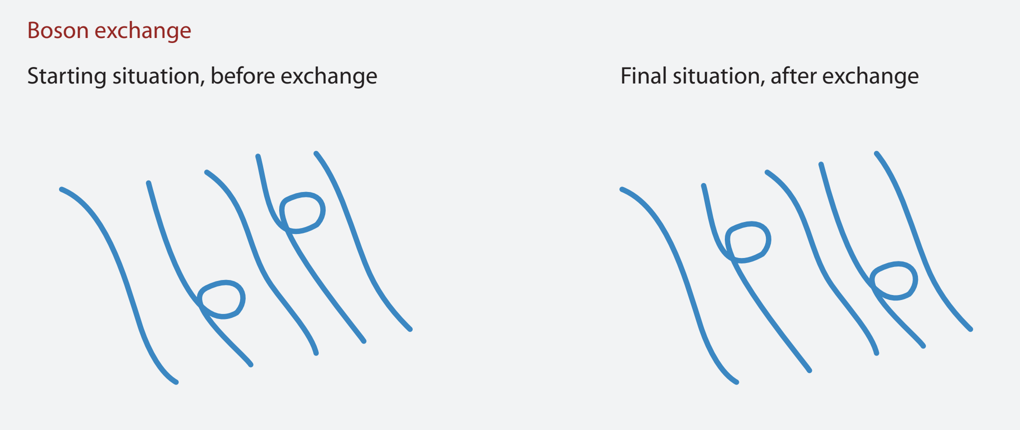


FIGURE 30 In the strand model, unknotted *boson* tangles can switch states without generating crossings, and thus without changing the sign of the phase.

elled with the help of a topology change of space or space-time. The various belt trick animations given above prove that this is not correct: spin 1/2 can be modelled in three dimensions in all its aspects. No topology change is needed. You might want to model the creation of a spin 1/2 particle–antiparticle pair as a final argument. Challenge 130 e

Bosons and spin

For tangles made of *one* strand – thus with two tails to the border – a rotation of the tangle core by 2π restores the original state. Such a tangle, shown in Figure 29, thus behaves like a spin 1 particle. The figure also shows the wave function that results from time averaging the crossings.

Bosons are particles whose combined state does not change phase when two particles are exchanged. We note directly that this is impossible with the tangle shown in Figure 29; the feat is only possible if the boson tangle is made of *unknotted* strands. Indeed, Vol. IV, page 115 for unknotted strands, the exchange process can easily switch the two deformations, as illustrated in Figure 30.

▷ Massive elementary particles thus can only be bosons if they also have an *unknotted* tangle in the tangle family that represents them.

The simplest strand model for each elementary boson – the photon, the W boson, the Z boson, the gluon and the Higgs boson – must thus be made of unknotted strands. We will deduce the precise tangles below, in the chapter on the particle spectrum. The tangle for the hypothetical graviton – also a boson, but in this case with spin 2 and invariant under core rotations by π – will be introduced in the chapter on general relativity. Page 312, page 329 Page 277

In summary, unknotted tangles realize the spin–statistics theorem for particles with integer spin: radiation particles, which have integer spin, are automatically bosons.

Spin and statistics

We just saw that fluctuating strands reproduce the spin–statistics theorem for fermions and for bosons, and thus for all elementary particles, if appropriate tangles are used. Apart from this fundamental result, the strand model also implies that no spins lower than $\hbar/2$ are possible, and that spin values are always an *integer multiple of* $\hbar/2$. All this matches observations. Challenge 131 e

In the strand model, temporal evolution and particle reactions *conserve* spin, because all interactions conserve the number of strands and tails. The details of the conservation will become clear later on. Again, the result agrees with observations.

The strand model thus explains the origin of permutation symmetry in nature: *permutation symmetry* of particles is due the possibility to exchange tangle cores of identical particles; and 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 161

In summary, the strand model reproduces the rotation, the spin and the exchange behaviour of elementary quantum particles – both fermions and bosons – in all its observed details. We now proceed to the next step: quantum mechanics of translational motion.

Tangle functions: blurred tangles

In the strand model, particle motion is due to the motion of tangle cores. But according to the fundamental principle, strands and tangles are not observable; only crossing switches are. To explore the relation between crossing switches and motion, 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 distance minimum, as shown in Figure 31. The sign of the orientation is defined by arbitrarily selecting one strand as the starting strand. The even larger arbitrariness in the definition of the phase will be of great importance later on, and lead to gauge invariance.

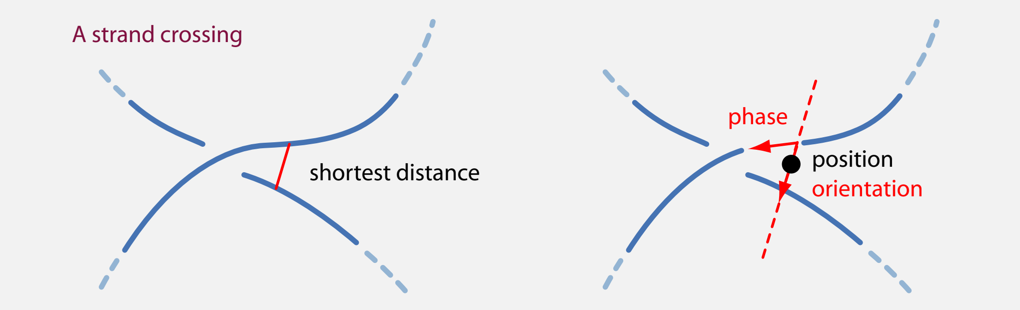


FIGURE 31 The definition of a crossing, its position, its orientation and its phase.

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 and *not* of the strands themselves).

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 32 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 its 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 – and in all known experiments – this will be 10^{-25} s or more; the typical averaging will thus be over a time interval with a value between 10^{-43} s, the Planck time, and around 10^{-25} s.

There are *two ways* to imagine tangle fluctuations and to deduce the short-time average

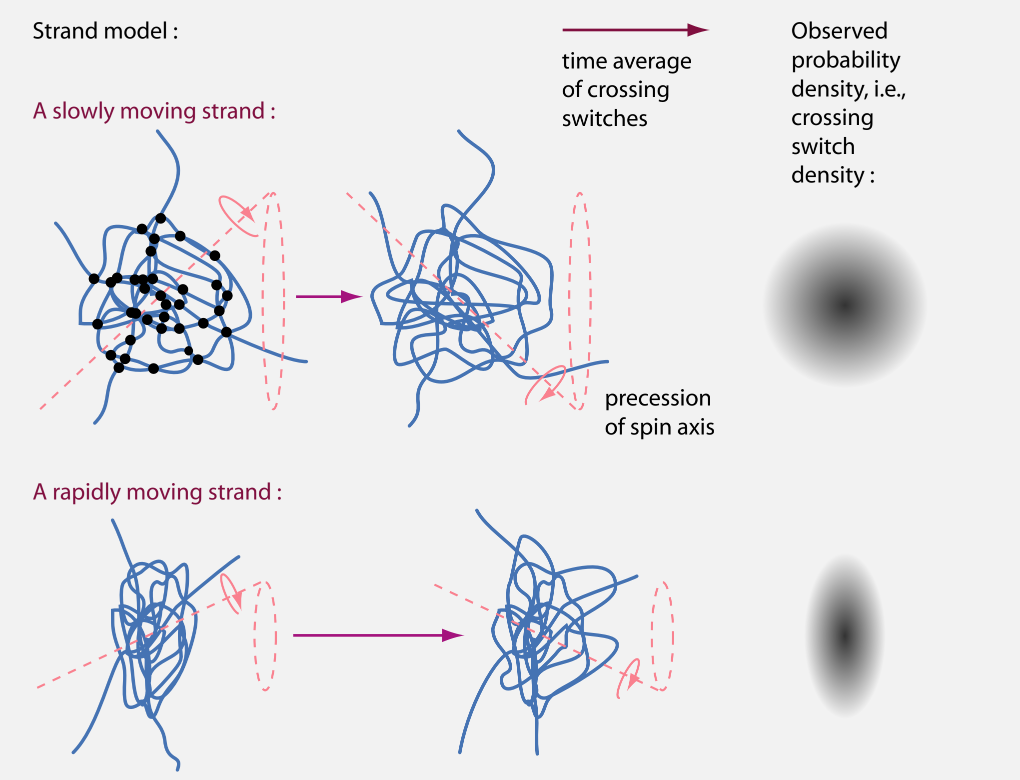


FIGURE 32 Some strand configurations, some of their short time fluctuations, and the corresponding probability density that results when averaging crossing switches over time. (The black dots are not completely drawn correctly.)

from a given tangle. The first, straightforward 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 picture* of the wave function and of quantum mechanics. The second, 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, in both cases the final result is that the tangle function is a cloud, i.e., a probability amplitude.

Challenge 132 e

DETAILS ON FLUCTUATIONS AND AVERAGES

In the strand model, the strand fluctuations of particle strands are a consequence of the embedding of all particles in a background which itself is made of fluctuating vacuum strands. Fluctuations randomly add detours to particle strands and 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 then can be said about the details of the averaging procedure for strand fluctuations?

The *fluctuations of the vacuum* are those strand fluctuations that lead to the definition of the background space. This definition 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. In particular, the fluctuations imply

▷ Flat vacuum has a tangle function that vanishes everywhere.

Challenge 133 e

The proof is an interesting exercise. The existence of a homogeneous and isotropic background space then implies conservation of energy, linear and angular momentum of particles travelling through it.

The *fluctuations of a tangle* lead, after averaging, to the tangle function, i.e., as we will see, to the wave function. The conservation of energy and momentum implies that the time average of the tangle fluctuations also conserves these quantities.

Page 291

Page 108, page 167

Therefore we can continue our discussion without yet knowing the precise details of the tangle fluctuations themselves. (We will provide these details below, in the section on general relativity.) Here we only require 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. Despite this limitation, such a circular description of nature must be *self-consistent*.

Page 201

We will also show below that the definition of the tangle function does *not* introduce hidden variables, even though first impression might suggest the opposite. In fact, it is possible to define something akin to a strand evolution equation. However, it does not deepen our understanding of the evolution equation of the wave function.

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, counting the crossings in a volume, and taking the square root. The crossing density – more precisely, its square root – 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 also 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 and thus not taken into consideration; 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.

Without the neglect of spin, and especially when the spin axis can *change* over space, the description of orientation and phase averages require more details; 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 approximation 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:

$$\psi(x, t) = R(x, t)\, \mathrm{e}^{i\alpha(x,t)/2} \; . \qquad (122)$$

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.

Challenge 134 s Here is a fun challenge: how is the shortest distance between the strands, for a crossing located at position x and t, related to the magnitude, i.e., the absolute value $R(x, t)$, of the wave function?

Page 201 We note that if *many* particles need to be described, the many-particle tangle function defines a separate crossing density for each particle tangle.

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 define addition and multiplication for tangles, and take the time average *after* the tangle operation is performed.

▷ Second, fun definition: The ***scalar multiplication*** $a\psi$ of a state ψ by a complex number $a = re^{i\delta}$ is formed by taking a tangle underlying the tangle function ψ, then rotating the tangle core by the angle 2δ, and finally 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 33. The above definition of scalar multiplication is only 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 scalars.

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 $\chi = a_1\psi_1 + a_2\psi_2$.

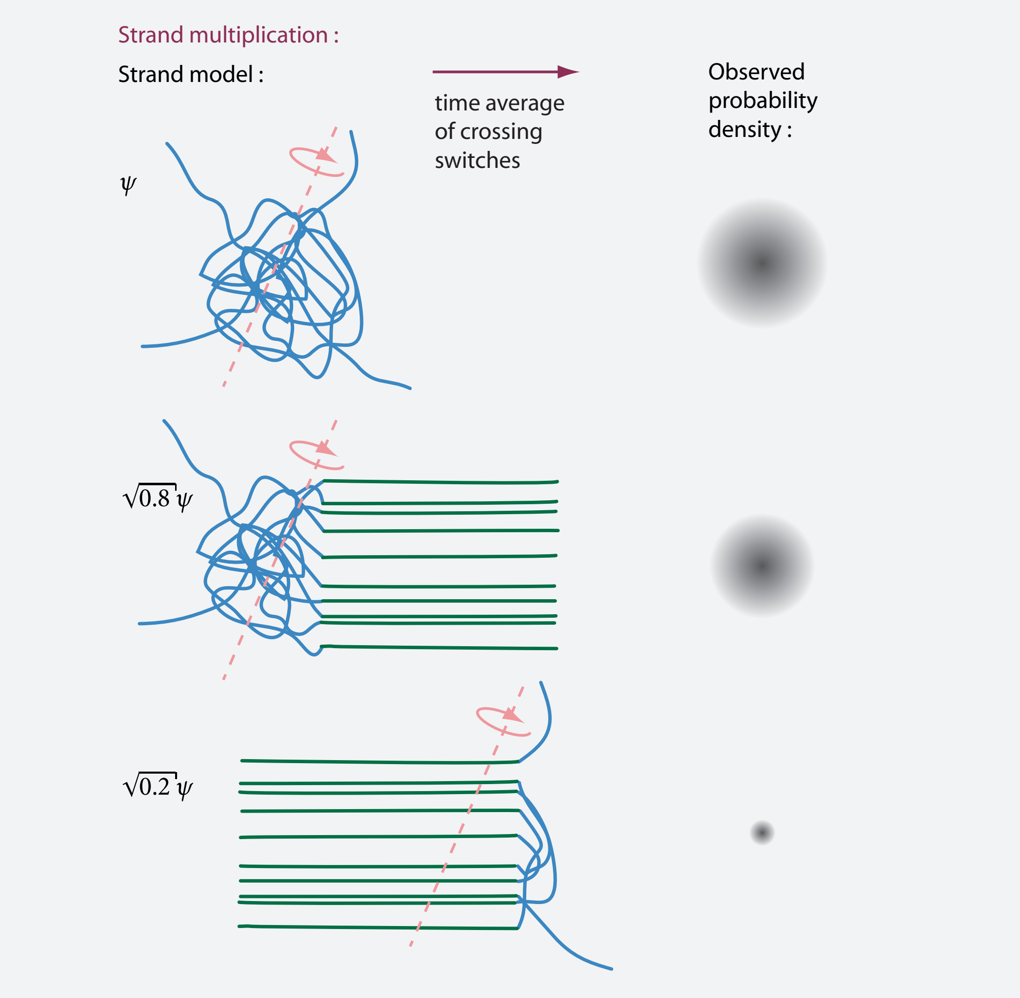


FIGURE 33 Scalar multiplication of localized tangles, visualizing the scalar multiplication of wave functions.

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 34. We note that despite the wording of the definition, no strand is actually cut or re-glued in the operation of addition.

The definition of linear combination requires that the final strand χ has the same topology and the same norm as each of the two strands ψ_1 and ψ_2 to be combined. Physically, this means that only states for the same particle can be added and that particle number is preserved; this automatically implements the so-called *superselection rules* of quantum theory. This 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. Page 190

The sum of two tangle functions is *unique*, for the same reasons given in the case of

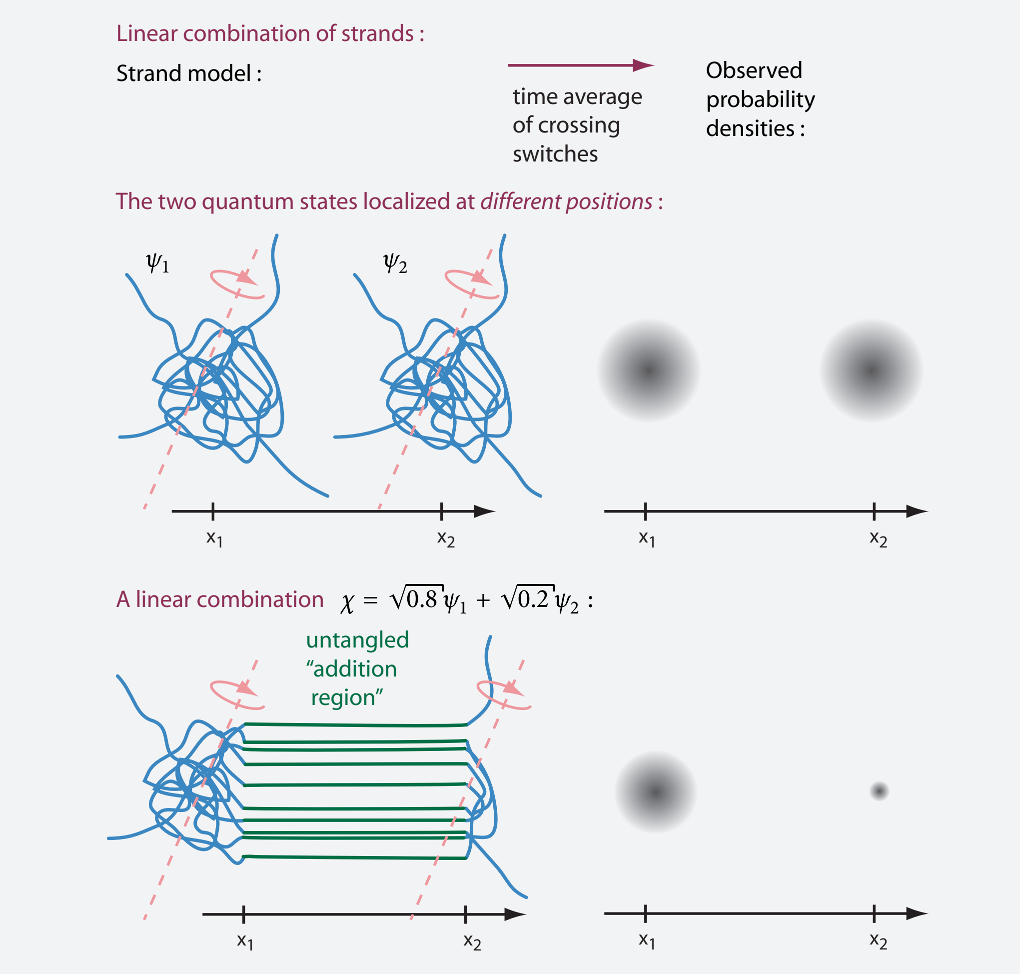


FIGURE 34 A linear combination of strands, in this case for two states representing a particle at two different position in space, visualizing the linear combination of wave functions.

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. 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 135 e

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 *crossing position 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 136 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 137 e

In summary, we have shown that tangle functions form a *Hilbert space*. The next steps are now obvious: 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 relativity, and show that they obey the Dirac equation.

DEDUCING THE SCHRÖDINGER EQUATION FROM TANGLES

The Schrödinger equation, like all evolution equations 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 this question, we will only need the fundamental principle that 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. Now, 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.* Rotating a tangle core leads to crossing switches in its tails. In the strand model, the kinetic energy E of a particle is thus due to the crossing switches formed in its tails. In other words, the kinetic energy E is related to the (effective) angular frequency ω of the core rotation by

$$E = \hbar\omega \,. \tag{123}$$

The local phase of the tangle function ψ changes with the rotation. This implies that

$$\omega = i\partial_t\psi \,. \tag{124}$$

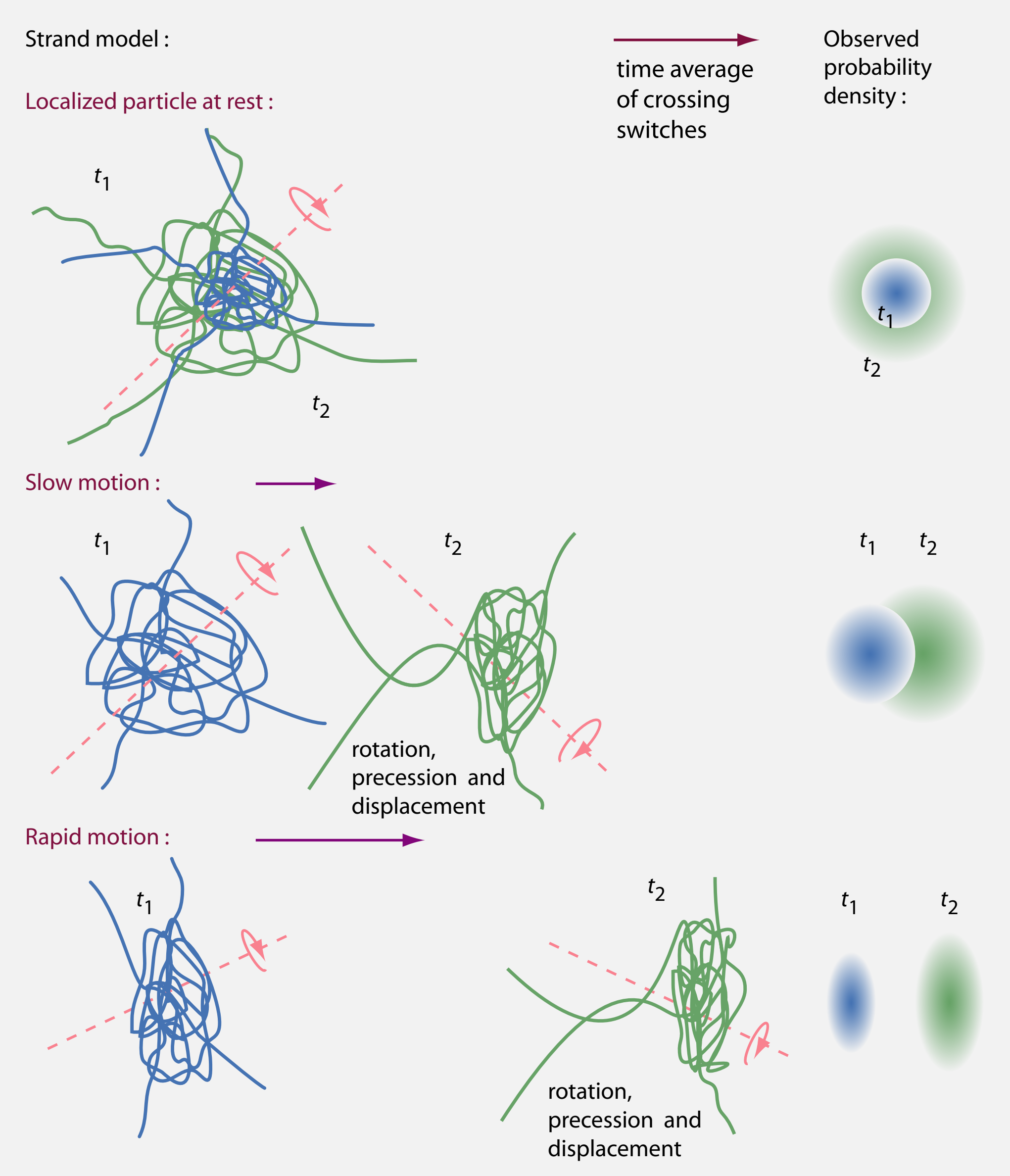


FIGURE 35 Examples of moving tangles of free particles.

We will need the relation shortly.

The linear motion of a tangle implies that it makes also sense to pay attention to 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 thus 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 (effective) wave number $k = 2\pi/\lambda$ of the core motion by

$$p = \hbar k . \tag{125}$$

The local phase of the tangle function ψ changes with the motion. This implies

$$k = -i\partial_x \psi . \tag{126}$$

This completes the description of matter wave functions without spin.

The belt trick for the fluctuating tails now has a fascinating consequence. To allow the belt trick also for high linear momentum, the more the momentum increases, the more the spin rotation axis has to align with the direction of motion. This is shown in Figure 35. 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. We thus get

$$E = \frac{p^2}{2m} \quad \text{and} \quad \omega = \frac{\hbar}{2m} k^2 . \tag{127}$$

This is dispersion relation for masses moving at velocities much smaller than the speed of light. The relation agrees with all experiments. The constant m is a proportionality factor that depends on the tangle core. We can now use the same argument that was used already by Schrödinger. Substituting the tangle relations in the dispersion relation, we get the evolution equation for the tangle function ψ given by

$$i\hbar\partial_t \psi = -\frac{\hbar^2}{2m} \partial_{xx} \psi . \tag{128}$$

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. In this way, we have also deduced, indirectly, Heisenberg's *indeterminacy relations.*

We have thus 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. We can sum up the situation in a few simple terms: *wave functions are blurred tangles.*

MASS FROM TANGLES

In quantum theory, particles spin while moving: the quantum phase rotates while a particle advances. The coupling between rotation and translation has a name: it is called the *mass* of a particle. We saw that the rotation is described by an average angular frequency ω, and the translational motion is described by a wave number k. The proportionality factor $m = \hbar k^2/2\omega = p^2/2E$ is thus a quantity that relates rotation frequency and wave number. In quantum theory,

▷ The (inertial) mass m describes the coupling between translation and rotation.

We note that a large mass value implies, for a given momentum value, both a slow translation and a slow rotation.

In the strand model, particle translation and rotation are modelled by the translation and rotation of the tangle core. Now, the strand model makes a point that goes beyond usual quantum theory. The strand model explains *why* core translation and rotation are coupled: When the core moves through the vacuum, the vacuum strands and the core effectively push against each other, due to their impenetrability. The result is a motion that resembles the motion of an asymmetrical body in a viscous fluid.

When an asymmetrical body is moved through a viscous fluid, it starts rotating. For example, this happens when a stone falls through water or honey. The rotation results Ref. 162 from the asymmetrical shape of the body. All the tangle cores of elementary particles are asymmetrical. The strand model thus predicts that tangle cores will rotate when they move through vacuum. In other terms, the strand model predicts

▷ Knotted tangles have mass.
▷ Unknotted tangles, such as those of photons, are predicted to be massless.

We also deduce that the more complicated a tangle is, the higher the mass value is.

In addition to the geometry effect due to the core, which is valid for massive bosons and fermions, the rotation of fermions is also influenced by the tails. The effective volume required by the belt trick will influence the coupling between translation and rotation. This effective volume will depend on the topology of the tangle core, and on the number of its tails. We again deduce that, for a given number of tails, a complicated core topology Page 353 implies a high mass value.

In other words, the strand model links the mass m of a particle to its tangle topology: *large tangle cores have large mass.* The strand model thus predicts

▷ Particle masses are *calculable* – if the tangle topology is known.

This is an exciting prospect! To sum up, the strand model predicts that experiments in viscous fluids can lead to a deeper understanding of the masses of elementary particles.

The tangle model also implies that the mass of elementary particles – thus of particles Challenge 138 e made of few strands – will be much smaller than the Planck mass. This is the first hint Page 355 that the strand model solves the so-called *mass hierarchy problem* of particle physics.

At this point, however, we are still in the dark about the precise origin of particle mass values. We do not know how to calculate them. Nevertheless, the missing steps are clear: first, we need to determine the tangle topology for each elementary particle; then we need to deduce their mass values, i.e., the relation between their rotation and translation. This is a central aim in the following.

An example of the issues that arise: How does the mass value depend on the number of strands in a tangle? How does mass depend on the type of tangle? Challenge 139 e

POTENTIALS

In quantum mechanics, interactions are described by potentials. An *electric potential* $V(x,t)$ changes the total energy of a particle with charge q at position x, since in quantum mechanics, electric potentials influence the rotation velocity of the wave function. As a result, with an electric potential, the left-hand side of the Schrödinger equation (128), the energy term, is changed from $\hbar\omega\psi(x,t)$ 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* $A(x,t)$ 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. We will deduce it in detail in the next chapter. Page 231

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:

$$(i\hbar\partial_t - qV)\psi = \frac{1}{2m}(-i\hbar\boldsymbol{\nabla} - q\boldsymbol{A})^2\psi\ . \qquad (129)$$

This equation is the simplest formulation of quantum theory. We saw in the fourth volume that it describes and explains the size of atoms and molecules, and thus of all objects around us; and we saw that it also explains the (relative) colours of all things. The equation also explains interference, tunnelling and decay.

In summary, a non-relativistic fluctuating tangle reproduces the full 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 following chapter, on gauge interactions. Page 221

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 the help of tangles.

Page 188 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 $\mathrm{e}^{i\beta}$. 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 19. Page 174

To deduce interference, we simply use the above definition of linear combinations of tangles. This leads to the result shown in Figure 36. 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 $\pi/4$.

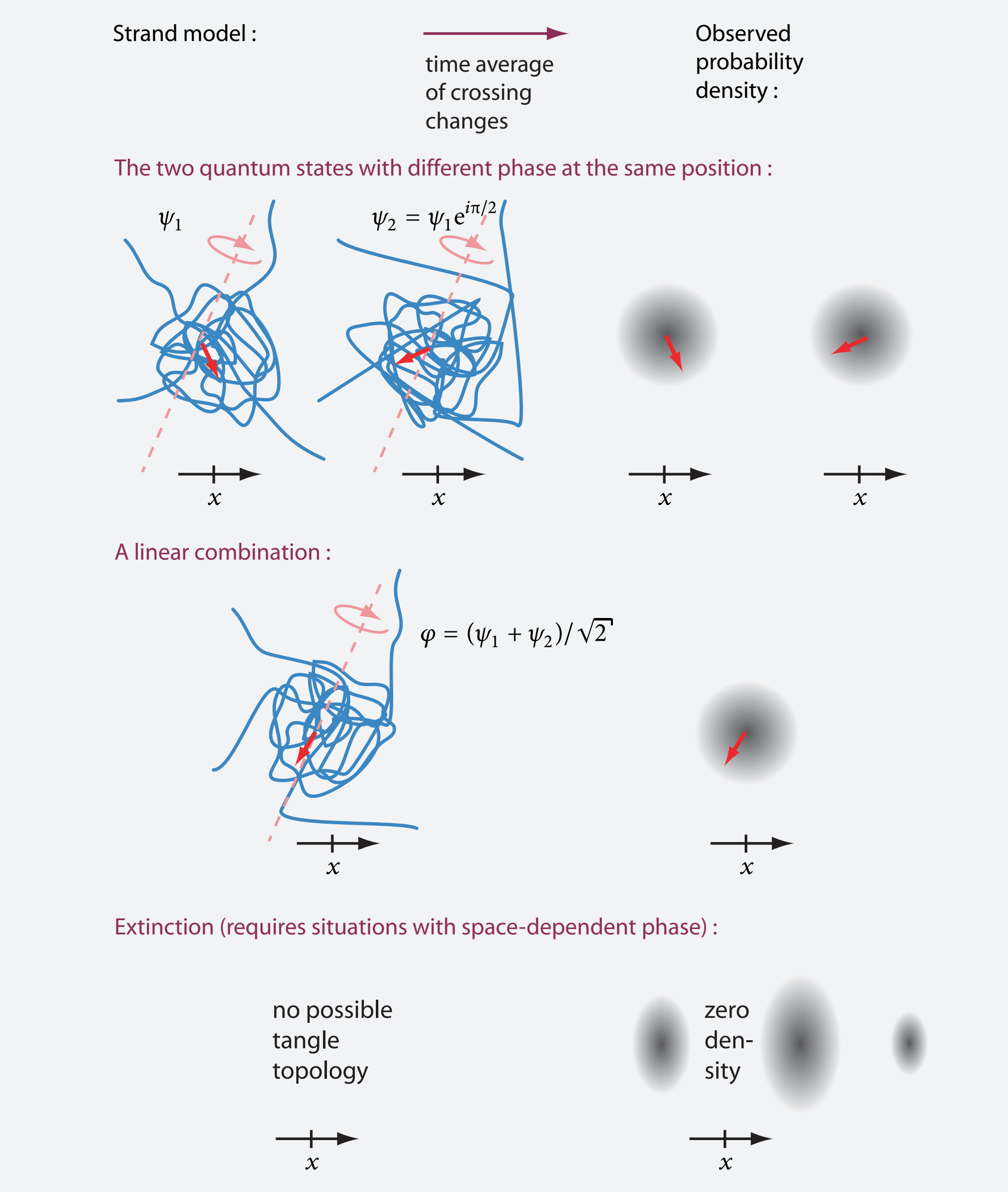


FIGURE 36 Interference: the linear combination of strands with different phase, but located at the same position.

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 particle tangle therefore must have vanishing crossing density in spatial regions where this operation is attempted. In short, particle tangles do 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 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. At 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 and γ. This yields a description of the tangle function as Ref. 163

$$\Psi(x,t) = \sqrt{\rho}\, \mathrm{e}^{i\alpha/2} \begin{pmatrix} \cos(\beta/2)\mathrm{e}^{i\gamma/2} \\ i\sin(\beta/2)\mathrm{e}^{-i\gamma/2} \end{pmatrix} , \qquad (130)$$

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 $\alpha(x,t)$, 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 $\beta(x,t)$ and $\gamma(x,t)$. 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 well-known physicist Paul Ehrenfest in analogy to 'vector' and 'tensor'; the English pronunciation is 'spinnor'.) For $\beta = \gamma = 0$, the previous wave function ψ is recovered.

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

$$\boldsymbol{\sigma} = \left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right) . \qquad (131)$$

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

$$i\hbar\partial_t \Psi = -\frac{\hbar^2}{2m}(\boldsymbol{\sigma}\nabla)^2\Psi . \qquad (132)$$

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 V and $\boldsymbol{A}$. A bit of algebra involving the spin operator then leads to the famous complete form of the Pauli equation

Challenge 140 s

$$(i\hbar\partial_t - qV)\Psi = \frac{1}{2m}(-i\hbar\nabla - q\boldsymbol{A})^2\Psi - \frac{q\hbar}{2m}\boldsymbol{\sigma B}\Psi , \qquad (133)$$

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–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.

Vol. IV, page 81

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.

Rotating arrows, path integrals and interference

Another simple way to visualize the equivalence between the strand model and the Pauli equation uses the formulation of quantum theory with path integrals. We recall that tangle tails are not observable, and that the tangle core defines the position and phase of the quantum particle. The motion of the core thus describes the 'path' of the particle. Different paths are due to different core motions.

The continuous rotation of the tangle core corresponds to Feynman's rotating little arrow in his famous popular book on QED. The different paths then correspond to different motions of the tangle core. The tangle model also reproduces the path integral formulation of quantum mechanics.

Ref. 164

Also interference can be visualized with strands. Because of its tails, a fermion tangle obeys spinor statistics and spinor rotation behaviour. This leads to the correct interference behaviour for spin 1/2 particles. Indeed, interference for fermions is visualized in Figure 37. The corresponding visualization for photon interference is given in Figure 38.

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.

Vol. IV, page 141

To put the issue into context, here is a short 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

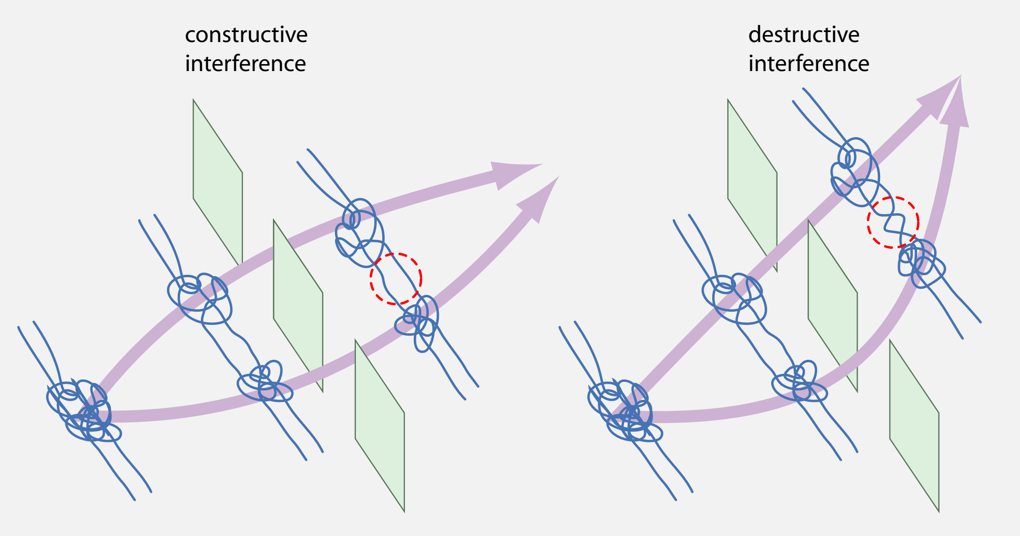


FIGURE 37 A fermion tangle passing a double slit: constructive interference (left) and destructive interference (right).

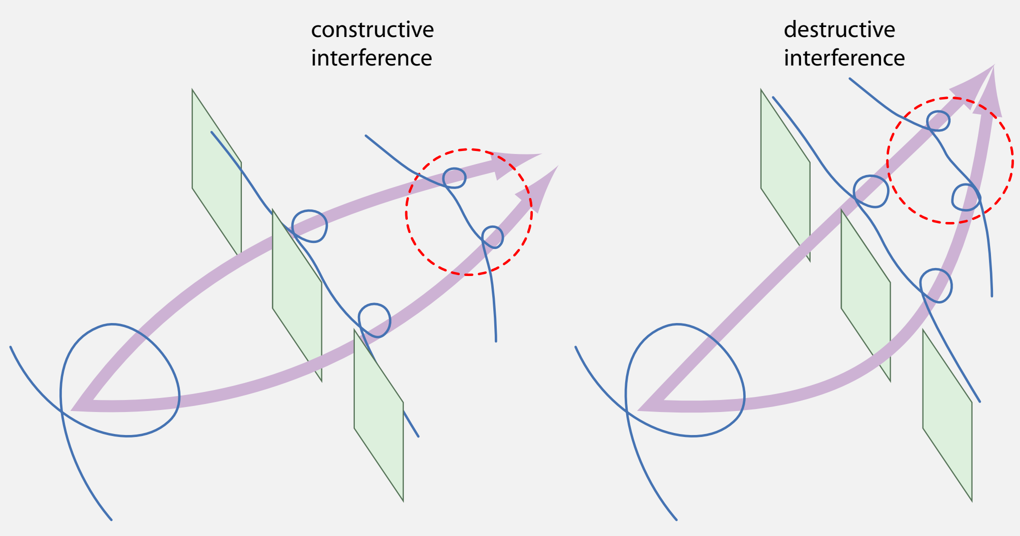


FIGURE 38 The double-slit experiment with photons: constructive interference (left) and destructive interference (right).

one or several baths. Thus, every measurement apparatus couples at least one bath to the system it measures. The coupling depends on and defines the observable to be measured by the apparatus. Every coupling of a bath to a *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.

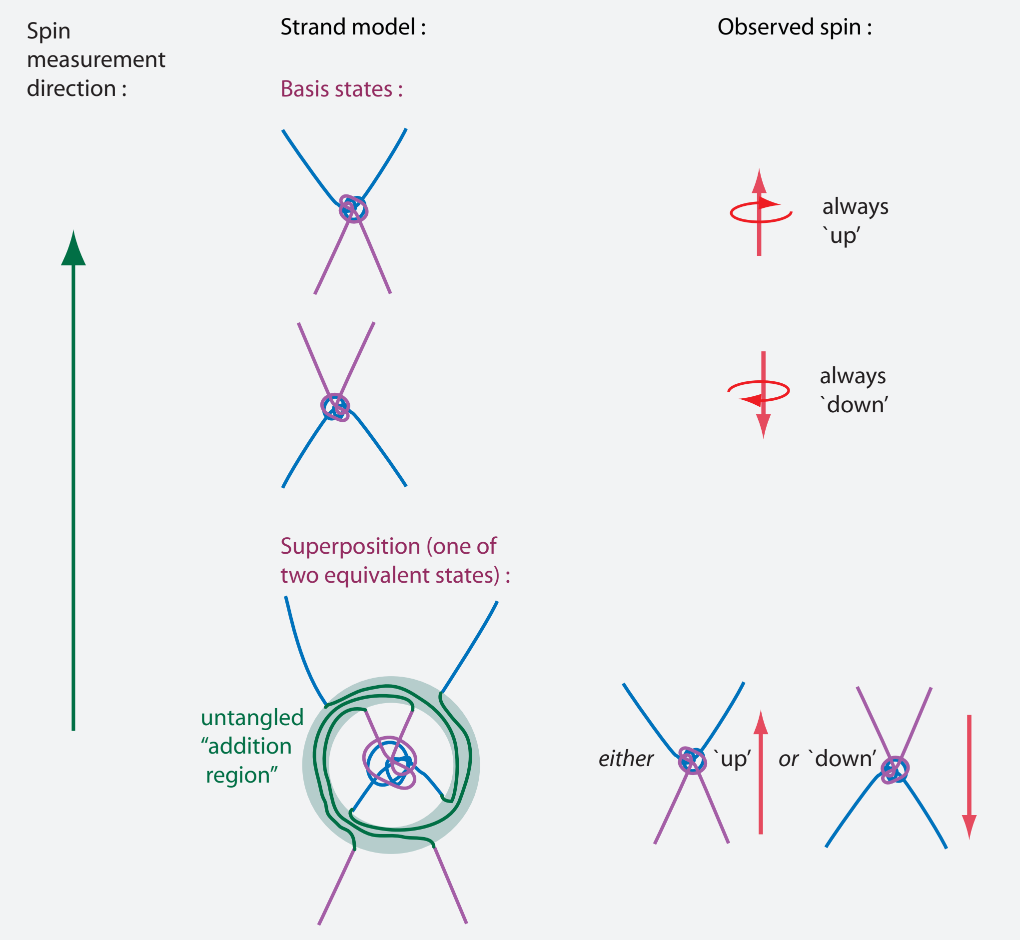


FIGURE 39 Measurement of a spin superposition: the addition region disappears either outwards or inwards.

- ▷ 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 39. 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 deduce that the probability of measuring a particular eigenstate will depend on the (weighed) volume that the eigenstate took up in 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 strands and at most 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 an eigenstate tangle.*

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–Specker theorem (for sufficiently high Hilbert-space dimensions). Is the strand model flawed? No. Ref. 165

We recall that strands are not observable. In particular, strand shapes are not physical observables and thus not physical variables either. Even if we tried promoting strand shapes to physical variables, the evolution of the strand shapes would only be observable 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. Thus

▷ The evolution of strand shapes and crossing switches is *contextual.*

Therefore, the strand model does not contradict the Kochen–Specker theorem.

In simple language, in quantum theory, hidden variables are not a problem if they are properties of the environment, and not of the quantum system itself. This is precisely the case for the strand model. For a quantum system, the strand model provides no hidden variables. In fact, for a quantum system, the strand model provides no variables beyond the usual ones from quantum theory. And as expected and required from any model that reproduces decoherence, the strand model leads to a contextual, 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.

Many-particle states and entanglement

In nature, the quantum states of two or more particles can be *entangled.* Entangled states are many-particle 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.

To describe entanglement, we first need to clarify the notion of many-particle state. In the strand model,

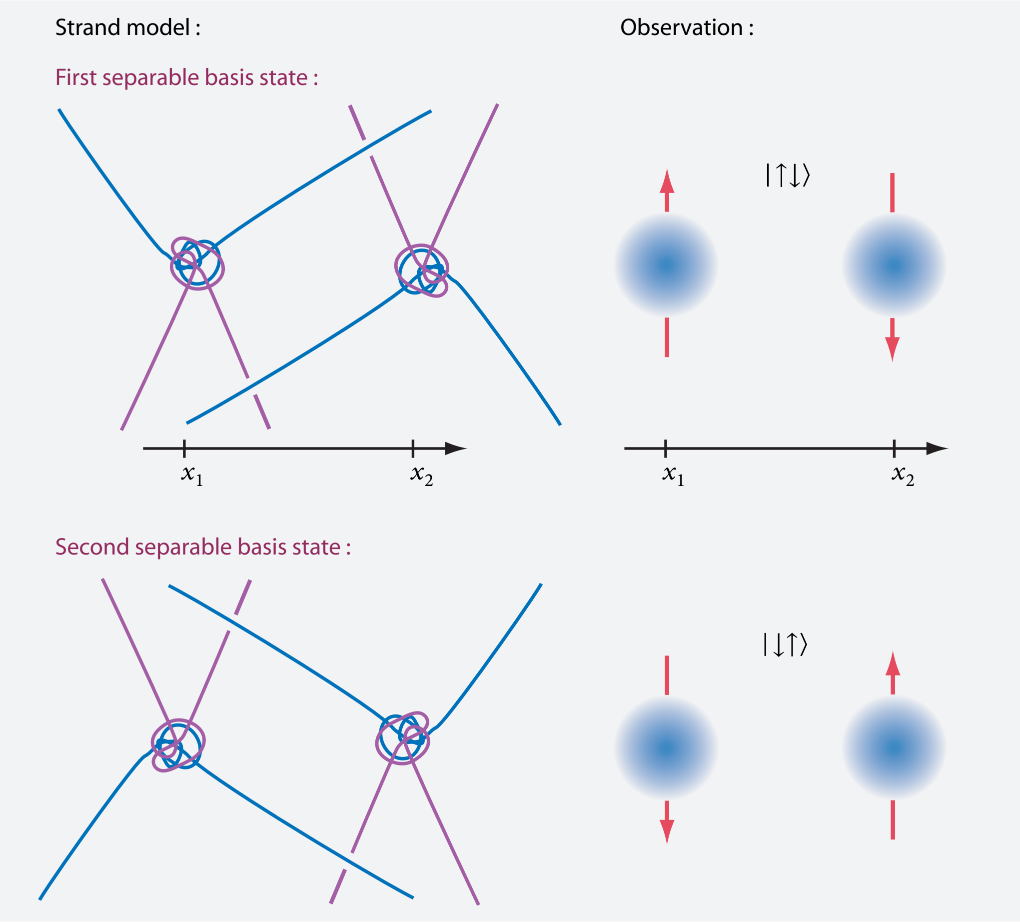


FIGURE 40 Two examples of two distant particles with spin in separable states: observation and strand model.

▷ A ***many-particle state*** is composed of several tangles.

In this way, an N-particle wave function defines N values at every point in space, one value for each particle. This is possible, because in the strand model, the strands of each particle tangle are *separate* from these of other particles.

Usually, a N-particle wave function is described by a single-valued function in $3N$ dimensions. It is less known that a single-valued N-particle wave function in $3N$ dimensions is mathematically equivalent to an N-valued wave function in three dimensions. Usually, N-valued functions are not discussed; we feel uneasy with the concept. But the strand model naturally defines N wave function values at each point in space: each particle has its own tangle, and each tangle yields, via short-term averaging, one complex value, with magnitude and phase, at each point in space. In this way, the strand model is able to describe N particles in just 3 dimensions.

In other words, the strand model does not describe N particles with 1 function in $3N$ dimensions; it describes many-particle states with N functions in 3 dimensions. In

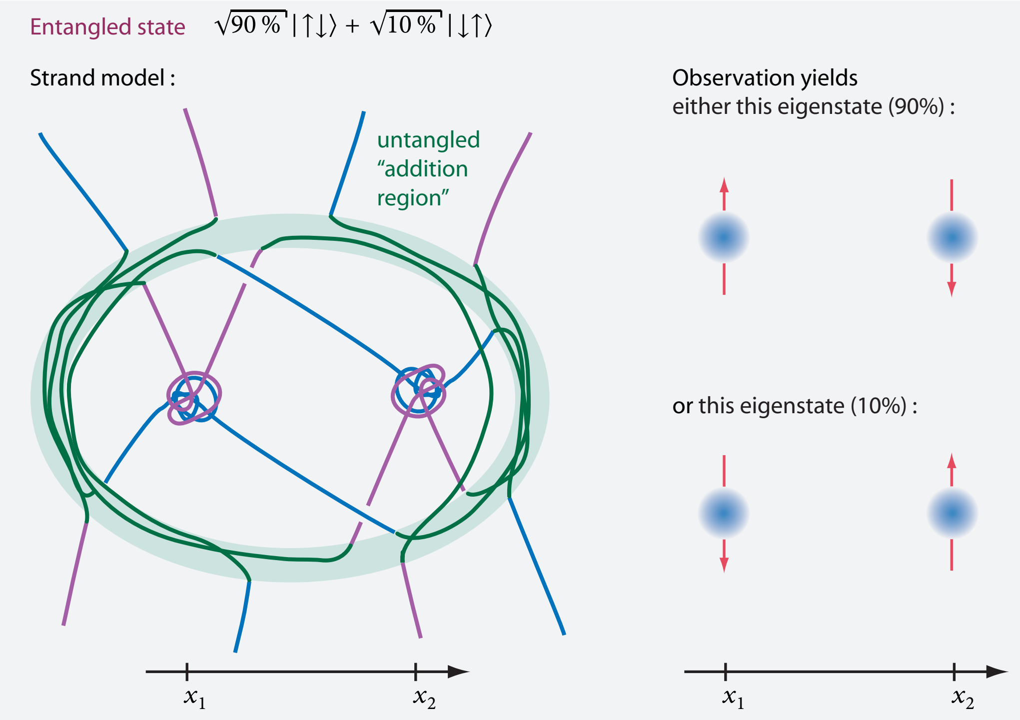


FIGURE 41 An entangled spin state of two distant particles.

this way, the strand model remains as close to everyday life as possible. Many incorrect statements on this issue are found in the research literature; many authors incorrectly claim the impossibility of many-particle quantum theory in 3 dimensions. Some authors even claim, in contrast to experiment, that it is impossible to visualize many-particle states in 3 dimensions. These arguments all fail to consider the possibility to define completely separate wave functions for each particle in three dimensions. (It must be said that this unusual possibility is hard to imagine if wave functions are described as continuous functions.) However, clear thinkers like Richard Feynman always pictured many-particle wave functions in 3 dimensions. Also in this domain, the strand model provides an underlying picture to Feynman's approach. This is another situation where the strand model eliminates incorrect thinking habits and supports the naive view of quantum theory.

Ref. 164

Now that we have defined many-particle states, we can also define entangled states.

▷ An ***entangled state*** is a non-separable superposition of separable many-particle states. State are separable when their tangles can be pulled away without their tails being entangled.

We will now show that the above definitions of superpositions and of measurements 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 thought experiment proposed by David Bohm. In the strand

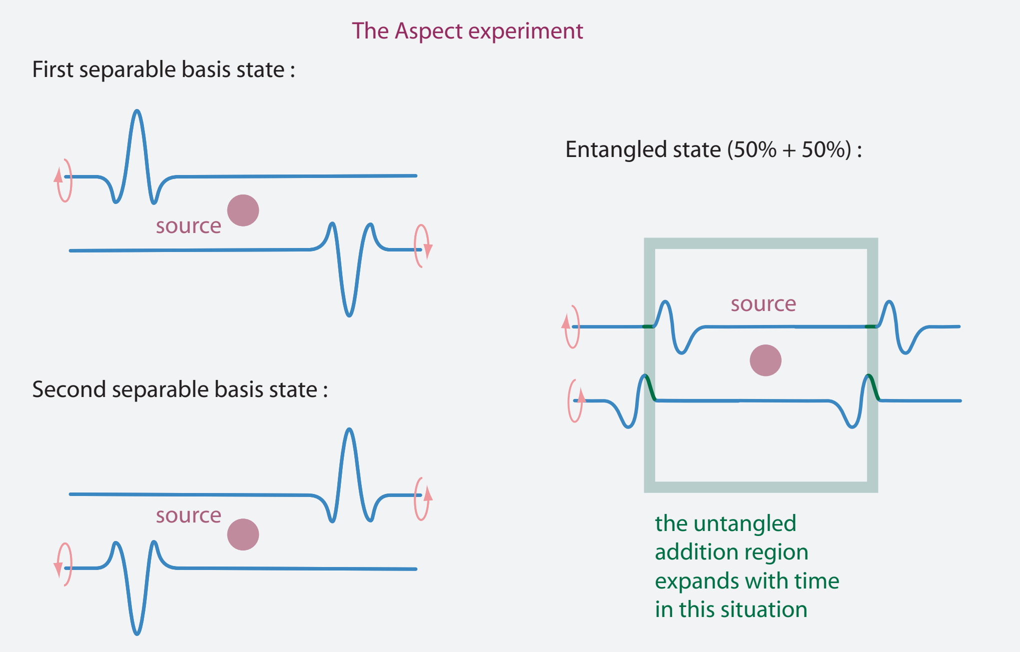


FIGURE 42 The basis states and an entangled state of two distant photons travelling in opposite directions, with total spin 0.

model, two distant particles with spin 1/2 in a *separable* state are modelled as two distant, separate tangles of identical topology. Figure 40 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 41. 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 experiment. 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. Here we use the strand model for the photon that we will deduce only later on. Figure 42 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 heli-

Page 226

city 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 Ref. 166 of many kilometres. Again, despite the extremely rapid collapse, no energy travels faster than light. And again, the strand model completely reproduces the observations.

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 argue that physical space must have three dimensions, as we will see now.

Deducing the dimensionality of physical space from first principles is an old and difficult problem. The difficulty is also due to the lack of alternative descriptions of nature. Page 167 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.

But how can we show that *physical space* – not the *background space* we need for thinking – is three-dimensional and must be so? 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, of particle exchange and all other observations confirm that space has three dimensions. Fermions only exist in three dimensions. In the strand model, the position and the orientation of a particle is intrinsically a three-dimensional quantity; physical space is thus three-dimensional, in all situations where it can be defined. (The only situations where this definition is impossible are horizons and the Planck scales.) 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.

Nature has three dimensions. The only way to predict this result is to show that no other number is possible. The number of dimensions of nature can only result from a self-consistency argument. And interestingly, the strand model produces such an argument.

In the strand model, knots and tangles are impossible to construct in physical spaces with dimensions *other* than three. Indeed, mathematicians can show that in four spatial dimensions, every knot and every tangle can be undone. (In this argument, time is not and does not count as a fourth spatial dimension, and strands are assumed to remain one-dimensional entities.) Worse, in the strand model, spin does not exist in spaces that have more or fewer than three dimensions. Also the vacuum and its quantum fluctuations do not exist in more than three dimensions. Moreover, in other dimensions it is impossible to formulate the fundamental principle. In short, the strand model of matter and of observers, be they animals, people or machines, is possible in three spatial dimensions only. No description of nature with a background or physical space of more or less than three dimensions is possible with strands. Conversely, constructing such a description would invalidate the strand model.

The same type of arguments can be collected for the one-dimensionality of physical 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. We have thus ticked off another of the millennium issues. We can thus continue with our adventure.

Challenge 141 e

Page 161

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 in a way that the corresponding wave function retains its norm, i.e., such that tangles retain their topology and their core 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 temporal 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 on the behaviour of operators. Already in 1987, Louis Kauffman had argued that the commutation relation for the momentum and position operators

Ref. 167

$$px - xp = \hbar i \tag{134}$$

is related to a crossing switch. The present section confirms 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 of a propagating state 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.

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 values are multiples of $\hbar$.

We note that these multiples, if averaged, 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 $W = \int \mathcal{L}\, \mathrm{d}V \mathrm{d}t = Tt$ 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*** $W = \int L\,\mathrm{d}t = \int\int \mathcal{L}\;\mathrm{d}V\mathrm{d}t$ 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

$$W_{\mathrm{if}} = \langle \psi_{\mathrm{i}} \,|\int \mathcal{L}\,\mathrm{d}t\,|\, \psi_{\mathrm{f}} \rangle = \langle \psi_{\mathrm{i}} \,|\int (\mathcal{T} - \mathcal{U})\,\mathrm{d}t\,|\, \psi_{\mathrm{f}} \rangle . \qquad (135)$$

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. With this connection, one can also show that the strand model implies Schwinger's quantum action principle. Challenge 142 e

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:

$$\mathcal{L} = \frac{\mathrm{i}\hbar}{2}(\overline{\psi}\,\partial_t \psi - \partial_t \overline{\psi}\,\psi) - \frac{\hbar^2}{2m}\nabla\overline{\psi}\nabla\psi . \qquad (136)$$

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.

We thus retain that the strand model explains the least action principle. It explains it in the following way: quantum evolution minimizes the number of crossing switches.

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 exactly the same for all inertial observers. In the strand model, the idea of flat vacuum as a set of 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. In the strand model, vacuum is *Lorentz-invariant*.

Many models of the vacuum, even fluctuating ones, have difficulties reproducing

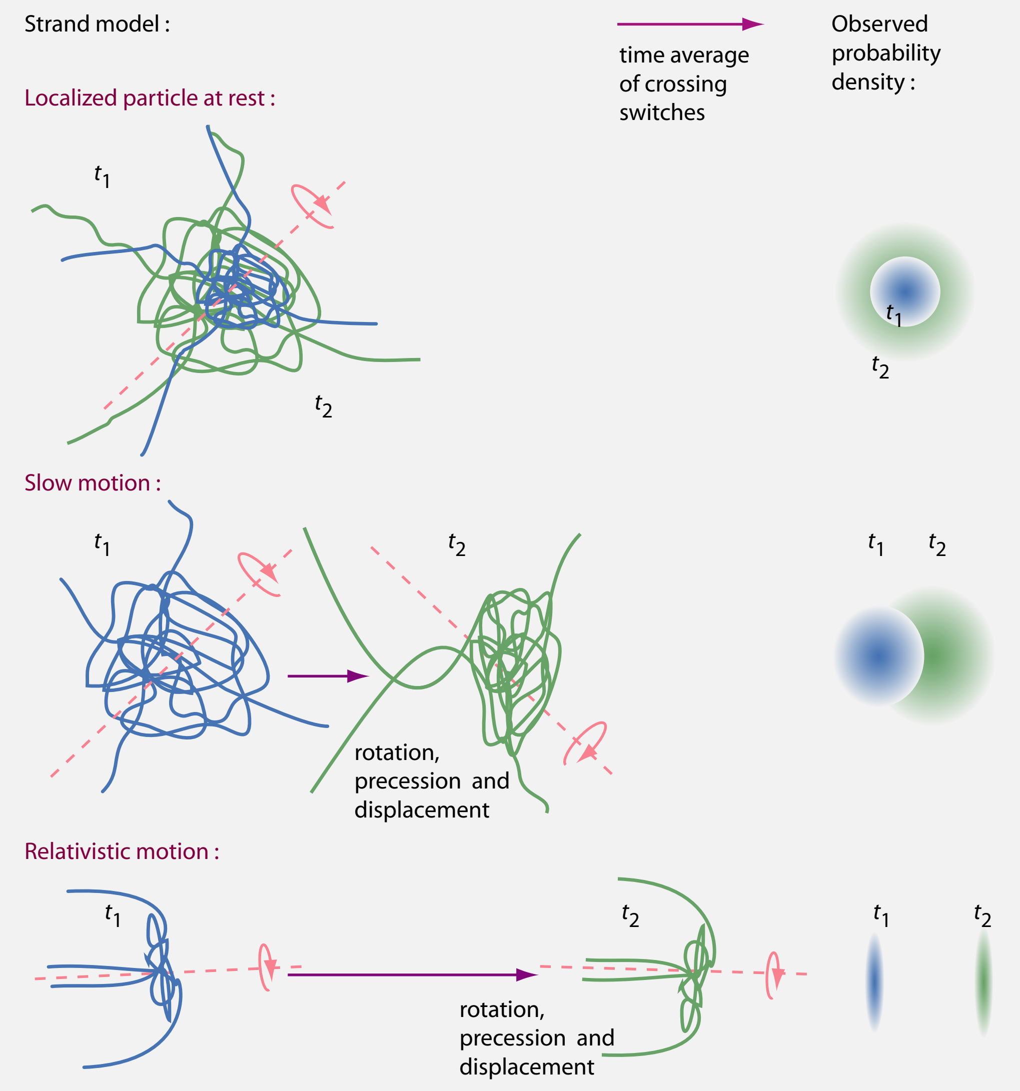


FIGURE 43 Tangles at rest, at low speed and at relativistic speed.

Lorentz invariance. The strand model differs, because the strands are not the observable entities; only their crossing switches are. This topological definition, together with the averaging of the fluctuations, makes the vacuum Lorentz-invariant.

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. In particular, there is no need for supersymmetry to explain the small energy density of the vacuum.

Special relativity: the invariant limit speed

Page 226 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 by a single, helically deformed unknotted strand, as shown in Figure 50. Therefore, we can define:

▷ The ***Planck speed*** c is the observed average speed of crossing switches due to photons.

Ref. 168 Because the definition uses crossing switches and a massless particle, the speed of light c is an *energy speed*. The speed of light c is also an average for long times. Indeed, as is well-known in quantum field theory, due to the indeterminacy relation, single photons can travel faster or slower than light, but the probability for large deviations is extremely low.

Page 349 The linear motion of a helically deformed photon strand through the vacuum strands is similar to the motion of a bottle opener through cork. It differs from the linear motion of a matter tangle through vacuum, which makes use of the belt trick. The belt trick slows fermions down, though the details are not simple, as we will discover below. In short, we find that matter tangles always move *more slowly than light*. The speed c is a *limit speed*.

In fact, we see that ultrarelativistic tangles move, as shown in Figure 43, almost like light. We thus find that matter can *almost* reach the speed of light. The speed c is indeed a *limit speed*.

Page 349 However, one problem remains open: how exactly do tangles move through the web that describes the vacuum? We will clarify this issue later on. In a few words, the motion of a photon requires that the strands of the surrounding space make room for it. This requires favourable fluctuations, thus a finite time. The motion process of photons thus makes it clear that the speed of light is *finite*.

The speed of light 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 mentioned properties of the speed of light – energy speed, limit speed, finite 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

$$E^2 = m^2c^4 + c^2p^2 \quad \text{or} \quad \hbar^2\omega^2 = m^2c^4 + c^2\hbar^2k^2 \ . \tag{137}$$

Page 353 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 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 Page 205 space-time deduced above, we have now definitely shown that flat vacuum has Poincaré Page 161 symmetry. This settles another issue from the millennium list.

Page 193 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 $E_0 = c^2 m$. Secondly, the spin precession is not independent of the particle speed any more; for relativistic particles, the spin lies close to the direction of motion. Both effects follow from the existence of a limit speed.

If we neglect spin, we can use the relativistic dispersion relation to deduce directly the well-known Klein–Gordon equation for the evolution of a wave function:

$$-\hbar^2 \partial_{tt}\psi = m^2 c^4 - c^2\hbar^2 \nabla^2 \psi \ . \tag{138}$$

In other words, the strand model implies that relativistic tangles follow the 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 Klein–Gordon equation assumes that spin effects are negligible. This approximation fails to describe most experiments. A precise description of relativistic elementary particles must include spin.

So far, we deduced the Schrödinger equation using the relation between phase and the quantum of action, using the non-relativistic energy–momentum relation, and neglecting spin. In the next step we deduced the Pauli equation by including the properties of spin 1/2. The following step was to deduce the 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) the effects of spin 1/2. Now we can reproduce this derivation because all three ingredients are reproduced by the strand model.

We first recall the derivation of the Dirac equation found in textbooks. The main observation about spin in the relativistic context is the existence of states of right-handed and of left-handed chirality: spin can precess in two opposite senses around the direction of momentum. In addition, 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 implicitly assumes only one, given sign for the chirality, even though it does not specify it. This simple description 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 energy–momentum relation, and Ref. 169 thus the Klein–Gordon equation. This requirement is known to be sufficient to deduce Ref. 170 the Dirac equation. One of the simplest derivations is due to Lerner; we summarize it

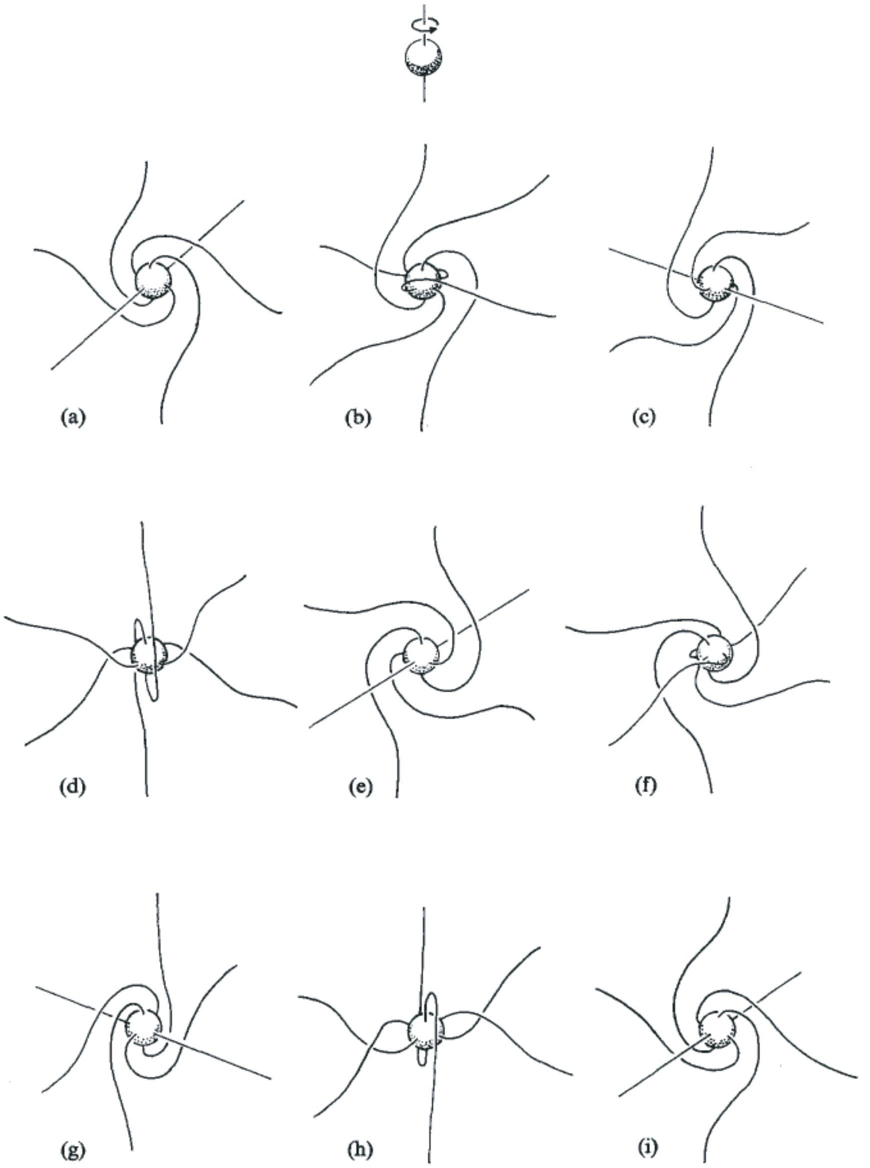


FIGURE 44 The belt trick for a rotating body with many tails, as used by Battey-Pratt and Racey to deduce the Dirac equation (© Springer Verlag, from Ref. 171).

here.

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 described by two complex numbers that depend on space and time. The 4-vector for probability and current becomes

$$J_\mu = u^\dagger \sigma_\mu u + v^\dagger \sigma_\mu v \,. \tag{139}$$

We now introduce the four-component spinor φ and the 4×4 spin matrices a_μ

$$\varphi = \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{and} \quad \alpha_\mu = \begin{pmatrix} \sigma_\mu & 0 \\ 0 & \overline{\sigma}_\mu \end{pmatrix} , \tag{140}$$

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

$$J_\mu = \varphi^\dagger \alpha_\mu \varphi . \tag{141}$$

The three requirements of current conservation, Lorentz invariance and linearity then yield the evolution equation Ref. 170

$$\mathrm{i}\hbar \partial^\mu (\alpha_\mu \varphi) + mc\gamma_5 \varphi = 0 . \tag{142}$$

This is the Dirac equation in the (less usual) spinorial representation.* The last term shows that mass mixes 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 above textbook derivation of the Dirac equation from usual quantum theory can be repeated and visualized also with the help of strands. There is no difference in arguments or results. The derivation with the help of strands was performed for the first time by Battey-Pratt and Racey, in 1980. Ref. 171 They explored a central object connected by unobservable strands (or 'tails') to the border of space, as shown in Figure 44. In their approach, the central object plus the tails correspond to a quantum particle. The central object is assumed to be continuously rotating, thus reproducing spin 1/2. They also assumed that only the central object is observable. (In the strand model, the central object becomes the tangle core.) Battey-Pratt and Racey then explored a relativistically moving object of either chirality. They showed that a description of such an object requires four complex fields. 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 textbook derivation just given.

We can thus say that the Dirac equation follows from the belt trick. We will visualize this connection in more detail in the next section. When the present author found this connection in 2008, Lou Kauffman pointed out the much earlier paper by Battey-Pratt and Racey. In fact, Paul Dirac was still alive when they found this connection, but unfortunately he did not answer their letter asking for comment.

In summary, tangles completely reproduce both the rotation and the linear motion of elementary fermions. Therefore, the strand model provides a simple view on the evolution equations of quantum theory. In the terms of the strand model, when spin is neglected, the Schrödinger equation describes the evolution of crossing density. For relativistic fermions, when the belt trick is included, the Dirac equation describes the

---

* The matrix γ_5 is defined here as

$$\gamma_5 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} , \tag{143}$$

where I is the 2×2 identity matrix.

evolution of crossing density. In fact, strands visualize these evolution equations in the most concrete way known so far.

Visualizing spinors and Dirac's equation using tangles

Despite its apparent complexity, the Dirac equation makes only a few statements: spin 1/2 particles are fermions, obey the relativistic energy–momentum relation, keep the quantum of action invariant, and thus behave like a wave. Each statement is visualized by the tangle model of fermions: tangles behave as spinors, the relativistic energy–momentum relation is built-in, the fundamental principle holds, and rotating tangle cores reproduce the evolution of the phase. Let us look at the details.

Given a particle tangle, the short-time fluctuations lead, after averaging of the crossings, to the wave function. The tangle model of fermions also 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 $\rho(x,t)$; physically, this is 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. It can also be seen as the axis around which the belt trick is performed.
- 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. 172

$$\varphi = \sqrt{\rho}\, \mathrm{e}^{i\delta}\, L(\boldsymbol{v})\, R(\alpha/2, \beta/2, \gamma/2) \;, \qquad (144)$$

where the product LR is an abbreviation for the boosted and rotated unit spinor and all parameters depend on space and time. This expression is equivalent to the description with four complex parameters used in most textbooks. In fact, 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, seventh phase. This is the phase that describes matter and anti-matter, that explains the appearance of the quantum of action $\hbar$, and that explains the fermion behaviour. Ref. 172

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 spinning relativistic quantum particles remains a simple extension

Ref. 164 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.

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. Ref. 172 The strand model provides a consistent visualization for these discussions.
2. The belt trick can be seen as the mechanism underlying the famous Zitterbewegung that is part of the Dirac equation. Ref. 173 The limitations in the observing the belt trick translate directly into the difficulties of observing the Zitterbewegung.
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. Ref. 174 Indeed, in the tangle model, particle rotation requires a regular application of the belt trick of Figure 19, Page 174 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.
6. The tangle model can be seen as a vindication of the stochastic quantization research programme; Ref. 175 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.

In short, rotating tangles are a correct underlying model for the propagation of fermions. And so far, tangles are also the only known correct model. *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), itself due to the fundamental principle.

And all three ingredients are reproduced by the strand model. We see that the apparent complexity of the Dirac equation hides its fundamental simplicity. The strand model reproduces the ingredients of the Dirac equation, reproduces the equation itself, and makes the simplicity manifest. In fact, we can say:

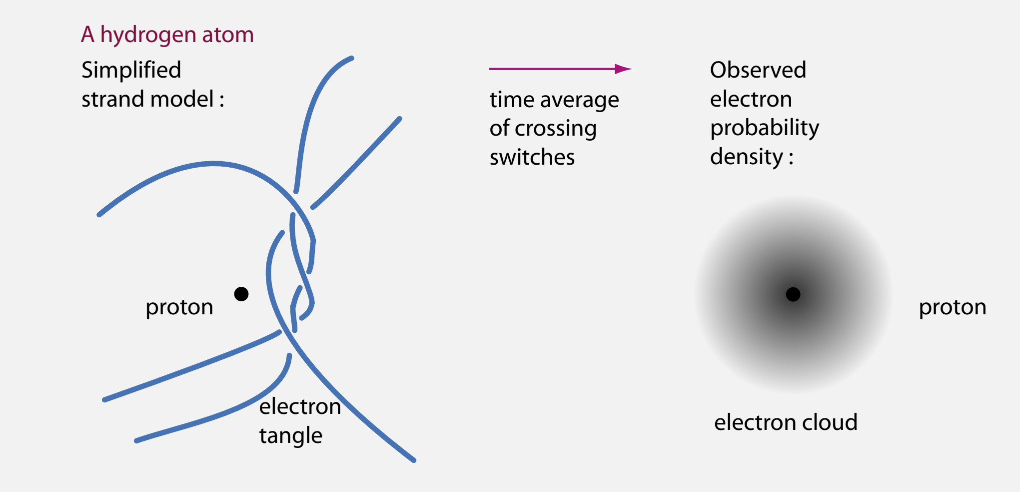


FIGURE 45 A simple, quantum-mechanical view of a hydrogen atom.

▷ The Dirac equation describes the relativistic infinitesimal belt trick or string trick.

The belt trick is fundamental for understanding the Dirac equation. In the strand model, core rotations vary along two dimensions – the rotation is described by two angles – and so does the belt trick. The resulting four combinations form the four components of the Dirac spinor and of the Dirac equation. Page 177

In summary, tangles can be used as a precise visualization and explanation of quantum physics. Wave functions, also those of fermions, are *blurred tangles* – with the detail that not the strands, but their crossings are blurred.

QUANTUM MECHANICS VS. QUANTUM FIELD THEORY

Quantum mechanics is the approximation to quantum physics in which fields are continuous and particles are immutable. In the strand model, quantum mechanics is thus the approximation in which a particle is described by a tangle with a shape that is *fixed* in time. This approximation allows us to derive the Dirac equation, the Klein–Gordon equation, the Proca equation, the Pauli equation and the Schrödinger equation. In this approximation, the strand model for the electron in a hydrogen atom is illustrated in Figure 45. This approximation already will allow us to deduce the existence of the three gauge interactions, as we will see in the next chapter.

In contrast, *quantum field theory* is the description in which fields are themselves described by bosons, and particles types can transform into each other. The strand model allows us to deduce the existence of all known gauge bosons, as shown in the next chapter. In the strand description of quantum field theory, particles are not tangles with a fixed shape of their core, but for each particle, the shape *varies*. This variation leads to gauge boson emission and absorption.

A FLASHBACK: SETTLING THREE PARADOXES OF GALILEAN PHYSICS

In all descriptions of physics, space and time are measured, explained and defined using matter. This occurs, for example, with the help of metre bars and clocks. On the other hand, matter is measured, explained and defined using space and time. This occurs, for example, by following a localized body over space and time. The circularity of the two definitions is at the basis of modern physics.

Page 108 As already mentioned above, the circularity is a natural consequence of the strand model. 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 counter-argument 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 from 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.

FUN CHALLENGES ABOUT QUANTUM THEORY

> “Urlaub ist die Fortsetzung des Familienlebens unter erschwerten Bedingungen.*”
> Dieter Hildebrandt

Are the definitions for the addition and multiplication of Schrödinger wave functions that were given above also valid for spinor tangle functions? Challenge 143 s

* *

The definition of tangle functions, or wave functions, did not take into account the crossings of the vacuum strands, but only those of the particle tangle. Why is this allowed? Challenge 144 e

* *

Modelling the measurement of action at the quantum level as the counting of full turns

---

* ‘Vacation is the continuation of family life under aggravated conditions.’ Dieter Hildebrandt (b. 1927 Bunzlau, d. 2013 Munich) was a cabaret artist, actor and author.

of a wheel is a well-known idea that is used by good teachers to take the mystery out of quantum physics. The strand model visualizes this idea by assigning the quantum of action ℏ to a full turn of one strand segment around another. Challenge 145 e

∗ ∗

Challenge 146 s Is any axiomatic system of quantum theory in contrast with the strand model?

∗ ∗

In the strand model, tangle energy is related to tangle core rotation. What is the difference between the angular frequency for tangles in the non-relativistic and in the relativistic case? Challenge 147 s

∗ ∗

Ref. 176 If you do not like the deduction of quantum mechanics given here, there is an alternative: you can deduce quantum mechanics in the way Schwinger did in his course, using the quantum action principle. Challenge 148 e

∗ ∗

Modern teaching of the Dirac equation replaces the spinor picture with the vector picture. Hrvoje Nikolić showed that the vector picture significantly simplifies the understanding of Lorentz covariance of the Dirac equation. Ref. 177 How does the vector picture clarify the relation between the belt trick and the Dirac equation? Challenge 149 r

∗ ∗

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 150 s

∗ ∗

A puzzle: Is the belt trick possible in a continuous and deformable medium – such as a sheet or a mattress – in which a coloured sphere is suspended? Is the belt trick possible with an *uncountably* infinite number of tails? Challenge 151 e

∗ ∗

Page 179 At first sight, the apheresis machine diagram of Figure 24 suggests that, using the belt trick, animals could grow and use wheels instead of legs, because rotating wheels could be supplied with blood and connected to nerves. Why did wheels not evolve nevertheless? Challenge 152 s

SUMMARY ON QUANTUM THEORY OF MATTER: EXPERIMENTAL PREDICTIONS

In this chapter, we used the fundamental principle – crossing switches define the quantum of action ℏ and the other Planck units – to deduce that particles are tangles of strands and that wave functions are time-averaged rotating tangles. In simple words,

- ▷ Both non-relativistic and relativistic wave functions are *blurred rotating tangles*.

More precisely, a wave function appears from the blurred crossings of a tangle. The components and phases of the wave function at a point in space are due to the orientation and phase of crossings at that point. We also deduced that blurred tangles obey the least action principle and the Dirac equation.

In other words, visualizing the quantum of action as a crossing switch implies quantum theory. The strand model confirms Bohr's statement: quantum theory is indeed a consequence of the quantum of action. Specifically, the strand model thus shows that all quantum effects are *consequences of extension* and *consequences of the three dimensions of space*. More precisely, all quantum effects are *due to tails*, the tails of the tangles that represent a quantum system. In particular, the strand model confirms that

- ▷ The Dirac equation is essentially the infinitesimal version of the belt trick (or string trick).

In other words, we have shown that strands reproduce the relativistic Lagrangian density $\mathcal{L}$ of charged, elementary, relativistic fermions in an external electromagnetic field $\boldsymbol{A}$

$$\mathcal{L} = \overline{\varphi}\,(i\hbar c \not{D} - c^2 m)\,\varphi\ , \qquad (145)$$

where

$$\not{D} = \gamma^\sigma \mathrm{D}_\sigma = \gamma^\sigma(\partial_\sigma - iqA_\sigma)\ . \qquad (146)$$

We thus conclude that *strands reproduce the quantum theory of matter*.

The strand model predicts deviations from the relativistic matter Lagrangian, and thus from the Dirac equation, *only* in three cases: first, when quantum aspects of electrodynamic field play a role, second, when nuclear interactions play a role, and third, when space curvature, i.e., strong gravity, plays a role. All this agrees with observation.

We will deduce the description of quantum electrodynamics and of the nuclear interactions in the next chapter. In the case of gravity, the strand model predicts that deviations from quantum theory occur exclusively when the energy–momentum of an elementary particle approaches the Planck value, i.e., for really strong gravity. Such deviations are not accessible to experiment at present. We will explore this situation in the subsequent chapter.

In addition, the strand model predicts that in nature, the Planck values for momentum and energy are limit values that cannot be exceeded by a quantum particle. All experiments agree with this prediction. Page 39

The deduction of quantum theory from strands given here is, at present, the *only* known microscopic explanation for quantum physics. So far, no other microscopic model, no different explanation nor any other Planck-scale deduction of quantum theory has been found. In particular, the extension of fundamental entities – together with observability limited to crossing switches – is the key to understanding quantum physics.

Page 161

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 the simplest part of the Lagrangian of quantum field theory, namely, the Lagrangian of free fermions, such as the electron, and that of fermions in continuous external fields. Therefore, for the next leg, we turn to the most important parts of the standard model Lagrangian that are missing: those due to gauge interactions.

Chapter 9

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 scales was still in the Page 18 dark. In the year 2000, it was known for several decades 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 Ref. 178 reason for this property was still unknown.

In this chapter we discover that fluctuating strands in three spatial dimensions explain the existence of precisely three gauge interactions, each with precisely the gauge symmetry that is observed. This is the first time ever that such an explanation is possible. Ref. 179 In other terms, we will deduce quantum field theory from strands. Indeed, strands provide a natural 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, because we shall discover a surprisingly simple result: *Gauge interactions and gauge symmetries are due to specific strand deformations.*

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. We thus start by exploring the non-gravitational interactions in the quantum domain.

Interactions and phase change

Experiments in the quantum domain show that interactions *change the phase* of wave functions. 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 a phase change. To explain this connection, we need to study the phase of tangle *cores* in more detail.

When we explored spin and its connection to the belt trick, we pictured the rotation Page 173 of the tangle core in the same way as the rotation of a belt buckle. We assumed that the core of the tangle rotates like a *rigid* object; the rotation is completed through the shape fluctuations of the tails only. Why did we assume this?

In Feynman's description of quantum theory, *free particles are advancing rotating arrows.* Ref. 164 In the strand model, *free* particle motion is modelled as the change of position of the tangle core and *spin* as the rotation of the core. We boldly assumed that the core

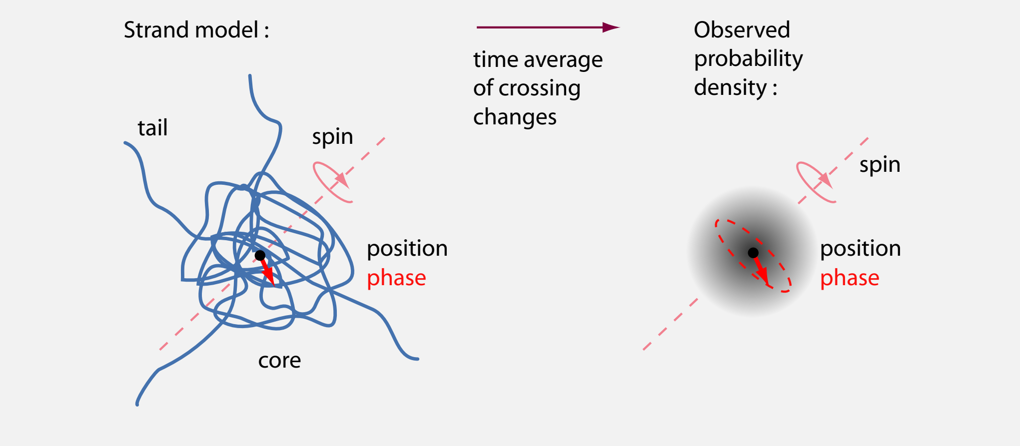


FIGURE 46 In the chapter on quantum theory, the phase was defined assuming a *rigidly rotating core*; this approximation was also used in the description of particle translation.

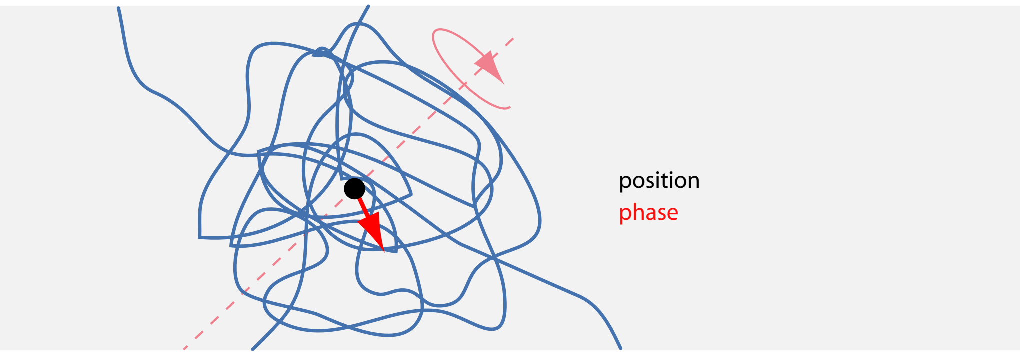


FIGURE 47 A magnified tangle core shows that the phase can also change due to *core deformations*; such core deformations lead to *gauge interactions*.

remained rigid, attached the phase arrow to it, and described spin as the rotation of the core with its attached arrow, as shown again in Figure 46. This bold simplification led us to the Dirac equation. In short, the assumption of a rigid core works.

However, we swept a problem under the rug: what happens if the core is *not* rigid? It turns out that the answer to this question automatically leads to the existence of gauge interactions. Now, we know from usual quantum theory that

▷ An ***interaction*** is a process that changes the phase of a wave function, but differs from a rotation.

In the strand model, shape deformations of tangle cores also lead to phase changes. In fact, we will discover that core deformations 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 free motion of matter tangles, as the chapter that focused on shape fluctuations of *tails*. Indeed, the belt trick completed the proof that

▷ ***Space-time symmetries*** are due to *tail* deformations.

All space-time symmetries – translation, rotation, boost, spin and particle exchange – are due to tail deformations; in such tail deformations, the tangle core is assumed to remain unchanged and rigid (in its own rest frame).

In contrast, the present chapter focuses on shape fluctuations in *tangle cores*.* We will discover that

▷ ***Gauge symmetries*** are due to *core* deformations.

Let us explore the tangle core in more detail. Figure 47 shows a magnified view of the core and its phase arrow. The phase of the core results from the phases of all its crossings. Thus, the figure makes it clear that the phase arrow will be sensitive to the shape fluctuations and deformations of the strand segments that make up the core.

In nature, any phase change of the wave function that is not due to a space-time symmetry is due to an interaction. For the strand model, this connection implies:

▷ When the phase of a core changes through *rigid orientation change*, we speak of ***core 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 exist? Secondly, how precisely is the phase – i.e., each arrow definition – influenced by core deformations? In particular, we have to check the answers and deductions with experiment.

The first question, on the classification of the core deformations, is less hard than Ref. 179 it might appear. The fundamental principle – events are crossing switches of strands – implies that deformations are observable only if they induce crossing switches. Other deformations do not have any physical effect. (Of course, certain deformations will have crossing switches for one observer and none for another. We will take this fact into consideration.) Already in 1926, the mathematician Kurt Reidemeister classified all those Ref. 181 tangle deformations that lead to crossing switches. The classification yields exactly three classes of deformations, today called the three *Reidemeister moves*. They are shown in Figure 48.

Ref. 180 * 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 Reidemeister moves on tangle cores represent the three gauge interactions, whereas the three Reidemeister moves on the vacuum represent (also) gravitational effects.

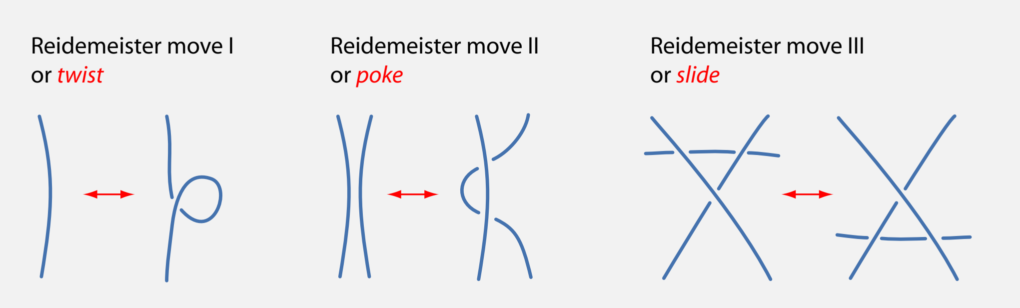


FIGURE 48 The Reidemeister moves: the three types of deformations that induce crossing switches – if the moves are properly defined in three dimensions.

- ▷ The ***first Reidemeister move***, or ***type I move***, or ***twist***, is the addition or removal of a twist in a strand.
- ▷ 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 type number of each Reidemeister move is also the number of involved strands. We will discover 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 situation of tangle cores.

The three Reidemeister moves turn out to be related to the three gauge interactions:

- ▷ The first Reidemeister move corresponds to *electromagnetism*.
- ▷ The second Reidemeister move corresponds to the *weak nuclear interaction*.
- ▷ The third Reidemeister move corresponds to the *strong nuclear interaction*.

We will prove this correspondence in the following.

For each Reidemeister move we will explore two types of core deformation processes: One deformation type are *core fluctuations*, which correspond, as we will see, to the emission and absorption of *virtual* interaction bosons. The other deformations are *externally induced core disturbances*, which correspond to the emission and absorption of *real* interaction bosons. As the first step, we show that both for fluctuations and for disturbances, the first Reidemeister move, the twist, is related to the electromagnetic interaction.

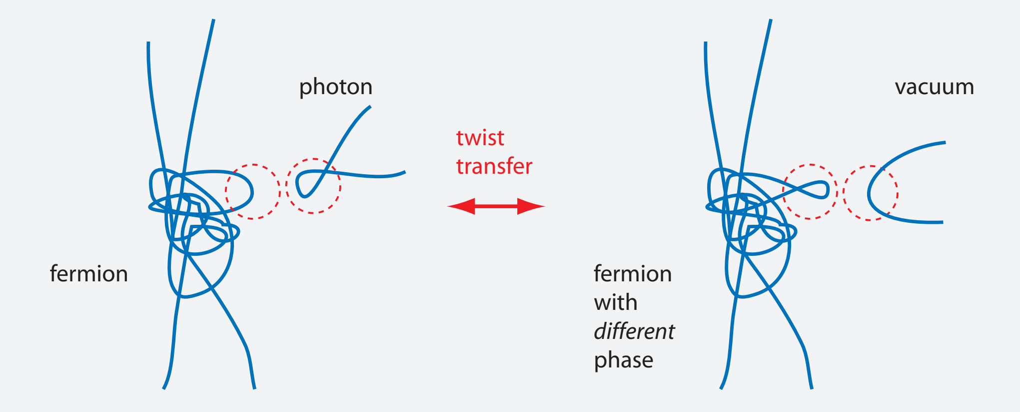


FIGURE 49 A single strand changes the rotation of a tangle: *twist transfer* is the basis of electromagnetism in the strand model. No strand is cut or reglued; the transfer occurs, statistically, through the excluded volume due to the impenetrability of strands.

ELECTRODYNAMICS AND THE FIRST REIDEMEISTER MOVE

Experiments show that all four fundamental interactions are described by potentials. Experiments also show that potentials change the phase, the rotation frequency and the wave number of wave functions. Experiments show that 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 elementary particles; neutral elementary 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. 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. Electromagnetism has a single fundamental Feynman diagram. The electromagnetic coupling constant at low energy, the so-called *fine structure constant*, is measured to be $\alpha = 1/137.035\,999\,139(31)$; its energy dependence is described by renormalization. Ref. 5

The previous paragraph contains everything known about the electromagnetic interaction. For example, Maxwell's field equations follow from Coulomb's inverse square relation, 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. In short, we now 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, massless spin 1 bosons such as the photon are made of a single strand. How can a single strand change the phase of a tangle? The answer is given in Figure 49: a *twisted loop* in a single strand will influence the rotation of a tangle because it changes the possible fluctuations of the tangle core. Due to the

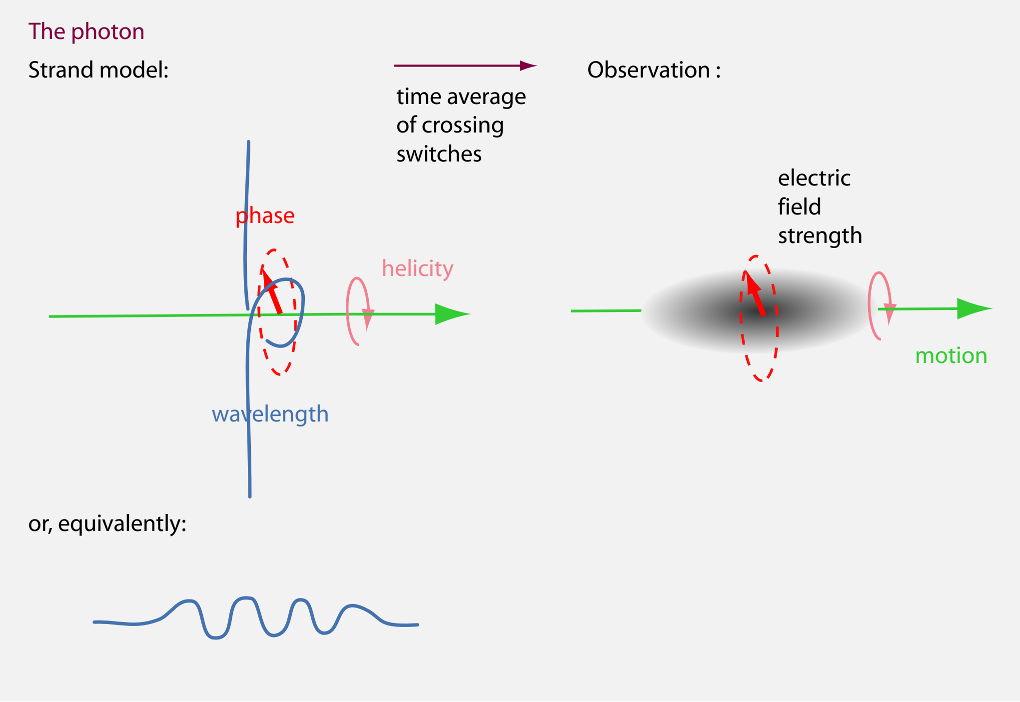


FIGURE 50 The photon in the strand model.

impenetrability of strands, an approaching twisted loop will sometimes transfer its twist to the tangle and thereby change its phase. The observed effect of an electromagnetic field on the phase of a charged fermion is the *time average* of all such twist transfers.

Page 173 Single strands represent bosons, as we saw above. Twisted loops are single strands and can have *two* twist senses, or two polarizations. Single, twisted and *unknotted* strands have no mass; in other words, twisted loops effectively move with the speed of light. And twisted loops, being curved, carry energy.

Approaching twisted loops will change the phase, i.e., the orientation of a matter tangle. Twisted loops correspond to a local rotation of a strand segment by π. But twists can be generalized to arbitrary angles. These generalized twists can be concatenated. Because they are described by a single angle, and because a double twist is equivalent to no twist at all, twists form a U(1) group. We show this in detail shortly. Page 231

In summary, twists behave like *photons* in all their properties. Therefore, the strand model suggests:

- ▹ A ***photon*** is a twisted strand. An illustration is given in Figure 50.
- ▹ The ***electromagnetic interaction*** is the transfer of twists, i.e., the transfer of first Reidemeister moves, between two particles, as shown in Figure 49.

The transfer of a twist from a single strand to a tangle core 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 of their impenetrability. No strand is ever cut or reglued.

CAN PHOTONS DECAY, DISAPPEAR OR BREAK UP?

The strand model of the photon, as shown in Figure 50, 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 might disappear. A helix might also disappear by a shape fluctuation.

A full image of a photon also includes the vacuum strands around it. In the strand model, the energy of the photon is localized in the configuration formed by the photon strand and the surrounding vacuum strands. In the strand model, energy is localized in regions of strand curvature. If the helical strands disappears, the surrounding vacuum strands are curved instead, or more strongly, and the energy is taken up by these surrounding strands. The net result is that the helix is transferred, permanently or for a short time, to another strand. In other terms, in the strand model, photons can also move by hopping from one strand to the next. Page 349

The only way in which a photon can disappear completely is by transferring its energy to a tangle. Such a process is called the *absorption* of a photon by a charged particle.

A single photon strand cannot break up into *several* photon strands of smaller helical diameters or of different rotation frequencies. Such a process is prevented by the fundamental principle. Challenge 153 e

In short, due to energy and to topological restrictions, the strand model prevents the decay, disappearance or splitting of photons, as long as no electric charge is involved. Linear and angular momentum conservation also lead to the same conclusion. Photons are *stable* particles in the strand model.

ELECTRIC CHARGE

Surrounded by a bath of photon strands, not all fermion tangles will change their phase. A tangle subject to randomly approaching virtual photons will feel a net effect over time only if it lacks some symmetry. In other words, only tangles that lack a certain symmetry will be electrically charged. Which symmetry will this be?

In a bath of photon strands, thus in a bath that induces random Reidemeister I moves, only *chiral* fermion tangles are expected to be influenced. In other terms:

▷ ***Electric charge*** is due to tangle chirality.

Conversely, we have:

▷ ***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 chirality can be right-handed or left-handed, there are positive and negative charges. Second, because strands are never cut or reglued in the strand model, chirality, and thus electric charge, is a *conserved quantity*. Third, chirality is only possible for tangles that are localized, and thus massive. Therefore, chiral tangles – charged

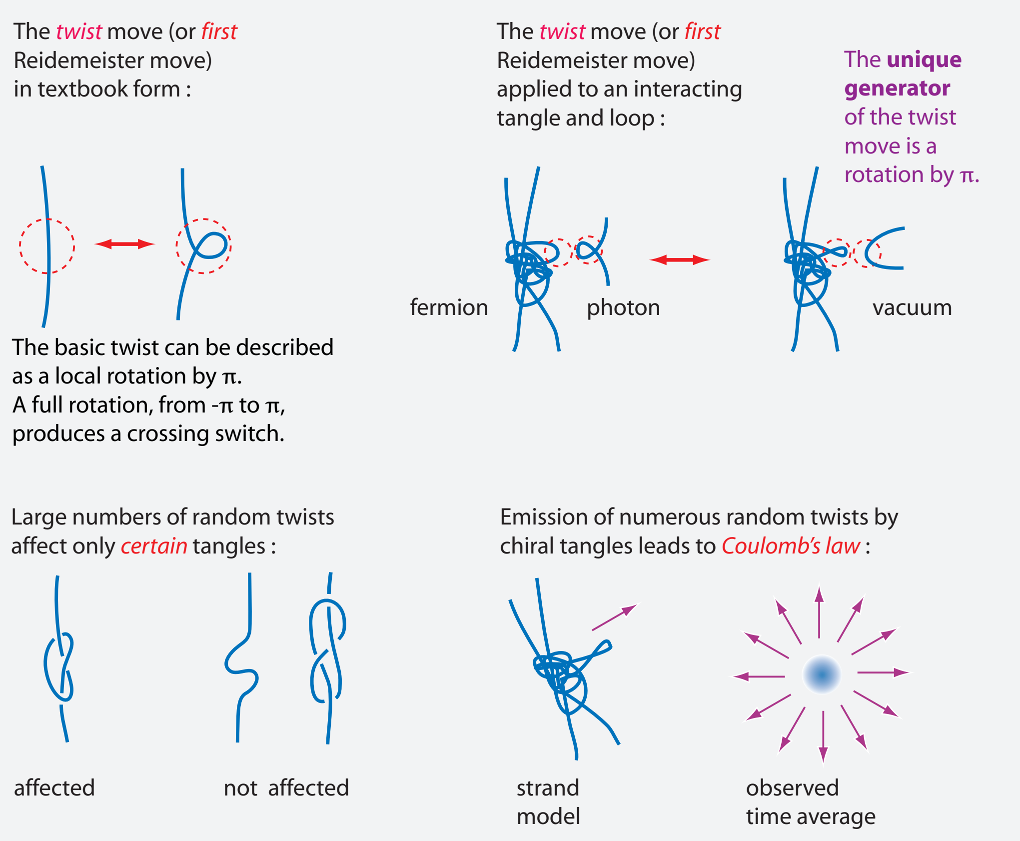


FIGURE 51 Electromagnetism in the strand model: the electromagnetic interaction, electric charge and Coulomb's inverse square relation.

particles – always move slower than light. Fourth, a chiral tangle at rest induces a twisted strand density around it that changes as $1/r^2$, as is illustrated in Figure 51. Finally, photons are uncharged; thus they are not influenced by other photons (to first order).

In short, all properties of electric charge found in nature are reproduced by the tangle model. We now check this in more detail.

Challenge: What knot invariant is electric charge?

Chirality explains the sign of electric charge, but not its magnitude in units of the elementary charge e. A full definition of electric charge must include this aspect.

Mathematicians defined various topological invariants for knot and tangles. *Topological invariants* are properties that are independent of the shape of the knot or tangle, but allow to distinguish knots or tangles that differ in the ways they are knotted. 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 417

A definition of electric charge q, proposed by Claus Ernst, is $q = c\,(n \bmod 2)$. Another option for the definition of charge is $q = w/3$. Equivalent definitions use the linking number. At this point of our exploration, the issue is open. We will come back to the detailed connection between charge, chirality and tangle topology later on. Page 383

ELECTRIC AND MAGNETIC FIELDS AND POTENTIALS

The definition of photons with twisted strands leads to the following definition.

▷ The ***electric field*** is the volume density of (oriented) crossings of twisted loops.

▷ The ***magnetic field*** is the flow density of (oriented) crossings of twisted loops.

▷ The ***electric potential*** is the density of twisted loops.

▷ The ***magnetic potential*** is the flow density of twisted loops.

The simplest way to check these definitions is to note that the random emission of twisted loops by electric charges yields Coulomb's inverse square relation: the force between two static spherical charges changes with inverse square of the distance. The strand model implies that in this case, *the crossing density is proportional to the square of the loop density*; in other words, the potential falls of as the inverse distance, and the electric field as the square distance.

The definition of the magnetic field simply follows from that of the electric field by changing to moving frame of reference. The two field definitions are illustrated in Figure 52.

We note that the electric field is defined almost in the same way as the wave function: both are oriented crossing densities. However, the electric field is defined with the crossing density of *twisted loops*, whereas the wave function is defined with the crossing density of *tangles*. The definitions differ only by the topology of the underlying strand structures. Page 183

In the strand model, energy, or action per time, is the number of crossing switches *per time*. The electromagnetic field energy per volume is thus given by the density of crossing switches *per time* that are due to twisted loops. Now, the strand model implies that *the crossing switch density per time is given by half the square of the crossing density plus half the square of the crossing density flow*. For twisted loops, we thus get that the energy density is half the square of the electric plus half the square of the magnetic field. Inserting the proportionality factors that lead from Planck units to SI units we get the well-known expression Challenge 154 e

$$\frac{E}{V} = \frac{\varepsilon_0}{2} E^2 + \frac{1}{2\mu_0} B^2 \, . \tag{147}$$

The strand model thus reproduces electromagnetic energy.

We note that in the strand model, the definition of the fields implies that there is no *magnetic charge* in nature. This agrees with observation.

The strand model predicts limit values to all observables. They always appear when strands are as closely packed as possible. This implies 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 $B_{\text{max}} = c^3/4Ge \approx$

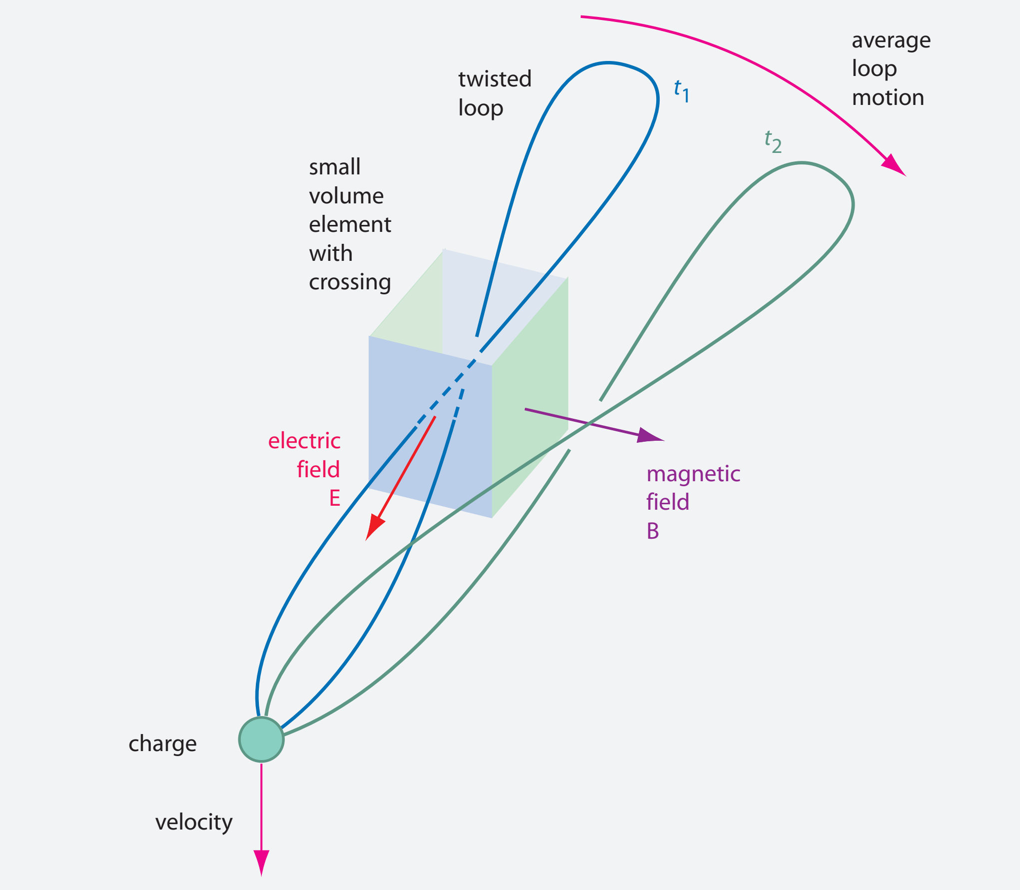


FIGURE 52 Moving twists allow us to define electric fields – as the density of twisted loop crossings – and magnetic fields – as the corresponding flow.

$6.3 \cdot 10^{53}$ T. All physical systems – including all astrophysical objects, such as gamma-ray bursters or quasars – are predicted to conform to this limit. Also this prediction agrees with observations.

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 F

$$\mathcal{L}_{\mathrm{EM}} = \frac{\varepsilon_0}{2} E^2 - \frac{1}{2\mu_0} B^2 = -\frac{1}{4\mu_0} F_{\mu\nu} F^{\mu\nu} \qquad (148)$$

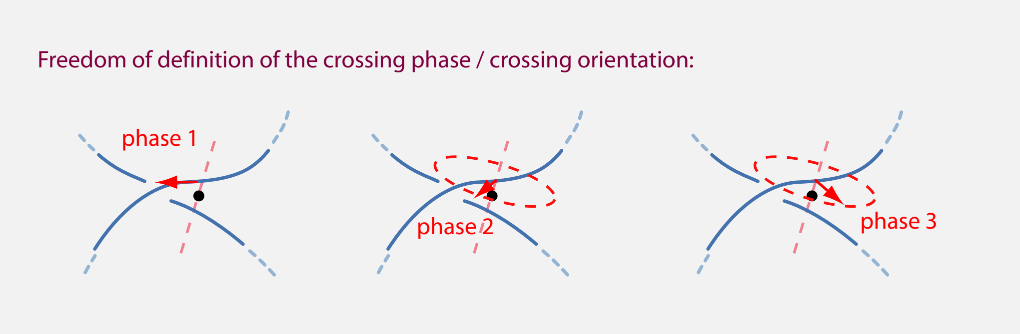


FIGURE 53 The definition of the phase or orientation of a single crossing is not unique: there is a freedom of choice.

where the electromagnetic field F is defined with the electromagnetic potential A as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \, . \tag{149}$$

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 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 electromagnetic potential A_μ is not uniquely defined: one says that 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 electromagnetic potential that have no effect on observations. In particular, gauge transformations leave unchanged all field intensities and field energies on the one hand and particle probabilities and particle energies on the other hand.

All these observations can be reproduced with strands. 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.

In the case of electrodynamics, the gauge freedom is a result of allowing phase choices that lie in a plane around the crossing orientation. (The other interactions follow from the other possible phase choices.) The phase choice can be different at every point in space. Changing the (local) phase definition is a (local) gauge transformation. Changing the phase definition for a single crossing implies changing the phase of wave functions and of the electromagnetic potentials. A schematic illustration of the choice of gauge is given in Figure 54.

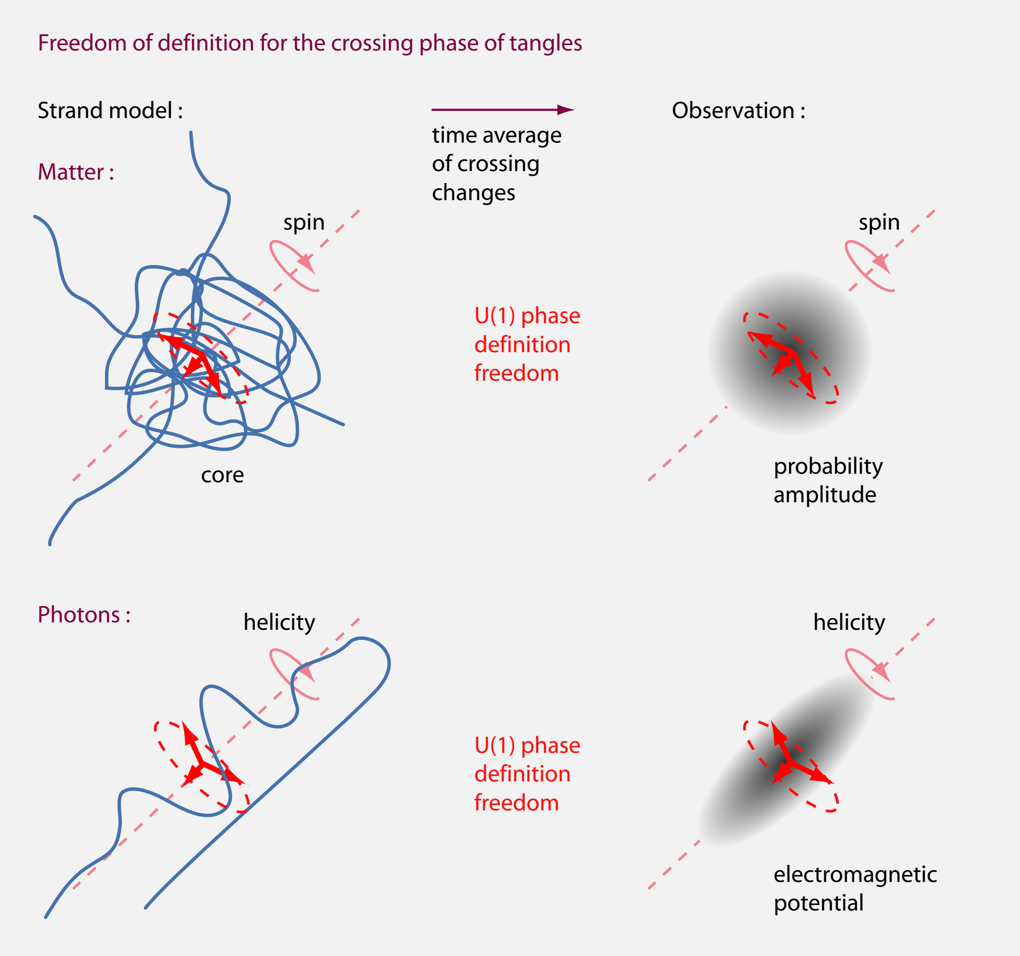
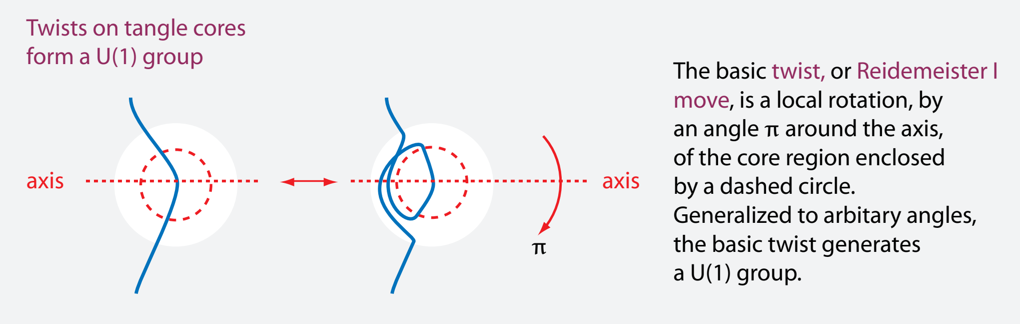


FIGURE 54 The freedom in definition of the phase of crossings leads to the gauge invariance of electrodynamics.



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 55 How the set of generalized twists – the set of all local rotations of a single strand segment around an axis – forms a U(1) gauge group.

We note that gauge transformations have no effect on the density or flow of crossings or crossing switches. In other words, gauge transformations leave electromagnetic field intensities and electromagnetic field energy invariant, as observed. Similarly, gauge transformations have no effect on the number of crossing switches of rotating tangles. A rotation by 4π does not change the phase, independently of which definition of arrow is chosen. Therefore, gauge transformations leave probability densities – and even observable phase differences – unchanged. This agrees with experiment.

A gauge transformation on a wave functions also implies a gauge transformation on the electrodynamic potential. The strand model thus implies that the two transformations are connected, as is observed. This connection is called *minimal coupling*. In short, minimal coupling is a consequence of the strand model.

U(1) GAUGE INTERACTIONS INDUCED BY TWISTS

There is only a small step from a gauge *choice* to a gauge *interaction*. We recall:

▷ A ***gauge interaction*** is a change of phase resulting from a strand deformation of the particle core.

In particular, electromagnetism results from the transfer of *twists*; twists are one of the three types of core deformations that lead to a crossing switch.

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 55. 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. The rotation axis is chosen as in shown by Figure 55. Generalized twists can be concatenated, and the identity twist – no local rotation at all – also exists. Generalized twists thus form a group. Furthermore, a generalized twist by 2π is equivalent to no twist at all, as is easily checked with a piece of rope:keeping the centre region is it disappears by pulling the ends, in contrast to a twist by π. These properties uniquely define the group U(1). In short, Figure 55 shows that generalized twists define the group U(1), which has the topology of a circle.

Challenge 155 e

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 full Lagrangian of QED.

THE LAGRANGIAN OF QED

Given the U(1) gauge invariance of observables, the Lagrangian of quantum electrodynamics, or QED, follows directly, because U(1) gauge invariance is equivalent to minimal coupling. We start from the Lagrangian density $\mathcal{L}$ of a *neutral*, *free*, and relativistic fermion in an electromagnetic field. It is given by

$$\mathcal{L} = \overline{\Psi}(i\hbar c \not\partial - c^2 m)\Psi - \frac{1}{4\mu_0} F_{\mu\nu}F^{\mu\nu} \, . \tag{150}$$

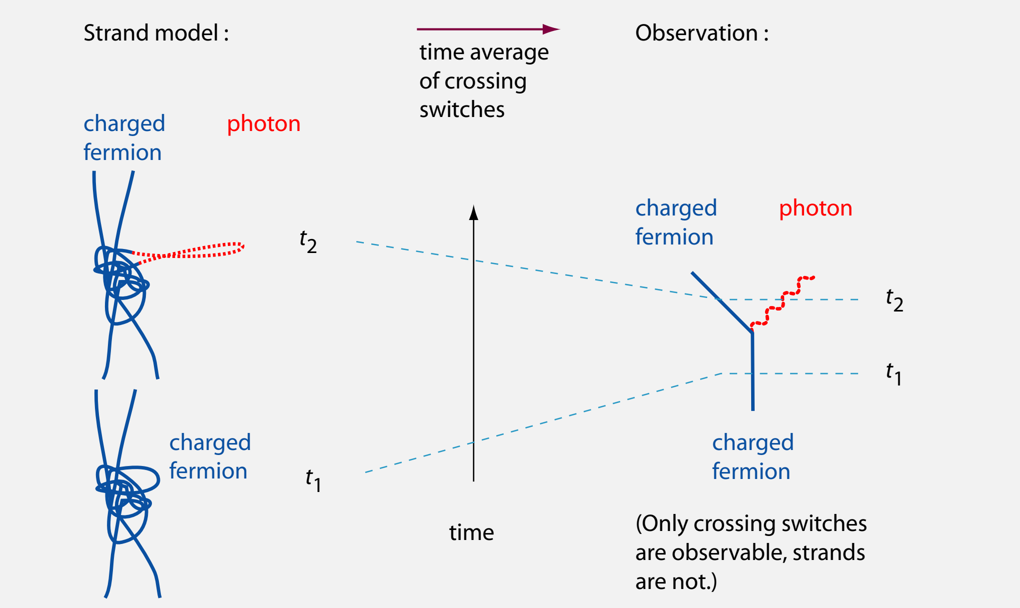


FIGURE 56 The fundamental Feynman diagram of QED and its tangle version.

Page 218 We deduced the fermion term in the chapter of quantum theory, and we deduced the electromagnetic term just now, from the properties of twisted loops.

As we have seen, the strand model implies minimal coupling. This changes the Lagrangian density for a *charged*, i.e., *interacting*, relativistic fermion in the electromagnetic field, into the Lagrangian density of QED:

$$\mathcal{L}_{\mathrm{QED}} = \overline{\Psi}(i\hbar c\not{D} - c^2 m)\Psi - \frac{1}{4\mu_0} F_{\mu\nu}F^{\mu\nu} \ . \qquad (151)$$

Here, $\not{D} = \gamma^\sigma \mathrm{D}_\sigma$ is the *gauge covariant derivative* that is defined through minimal coupling to the charge q:

$$\mathrm{D}_\sigma = \partial_\sigma - iqA_\sigma \ . \qquad (152)$$

Minimal coupling implies that the Lagrangian density of QED is invariant under U(1) gauge transformations. We will discuss the details of the charge q later on. Page 376

We have thus recovered the Lagrangian density of quantum electrodynamics from strands. Strands thus reproduce the most precisely tested theory of physics.

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 56. Important Feynman diagrams are shown on the left-hand sides of Figure 57 and of Figure 58.

In the strand model, the fundamental Feynman diagram can be visualized directly

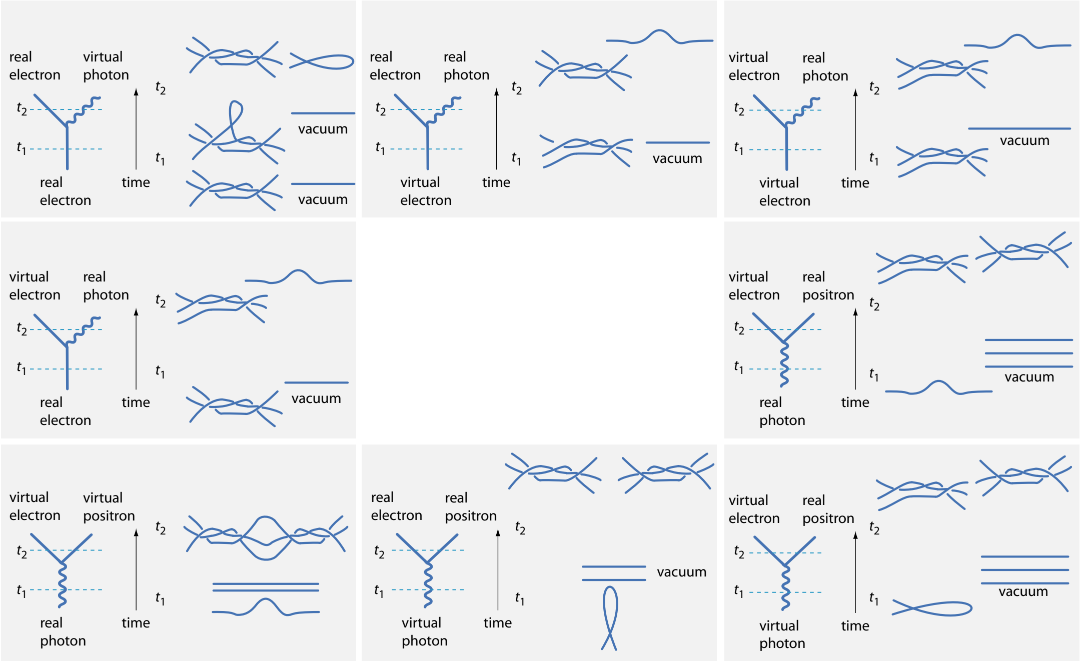


FIGURE 57 The different variations of the fundamental Feynman diagram of QED and their tangle versions.

Page 225

Page 312

in terms of strands, as shown on the left-hand side of Figure 56. 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 58.

In quantum field theory, Lagrangians must not only be Lorentz and gauge invariant, but must also be 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 in the limit 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 g-factors of charged particles, as shown in the next section. In particular, the strand model for QED, as shown in Figure 56, implies that higher order QED diagrams are simple *strand deformations* of lower order diagrams. Taking statistical averages of strand deformations up to a given number of crossings thus allows us to calculate QED effects up to a given order in the coupling. The strand model thus suggests that non-perturbative calculations are possible in QED. However, we do not pursue this topic in the present text.

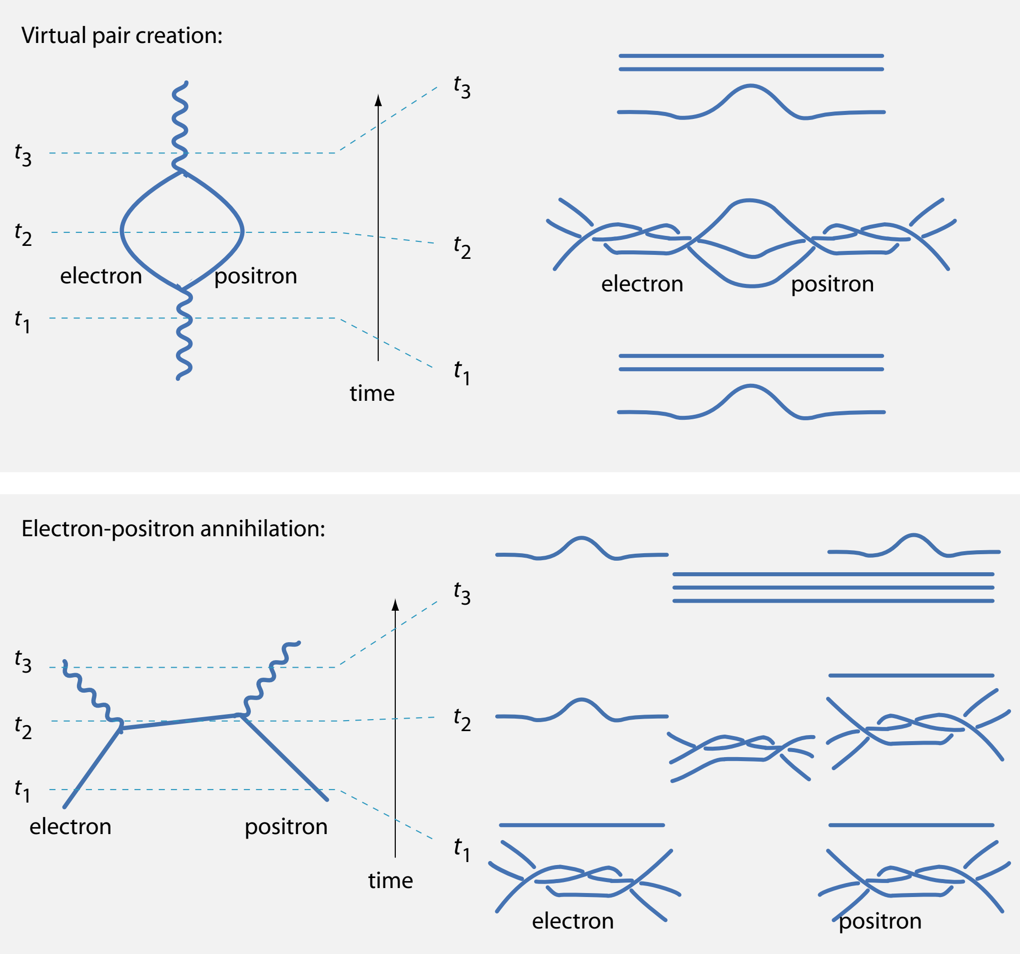


FIGURE 58 Some Feynman diagrams of QED with their tangle versions.

For precise non-perturbative calculations, the effective diameter of the strands must be taken into account. The diameter eliminates the Landau pole and all ultraviolet divergences of QED. In the strand model, the vacuum energy of the electromagnetic field is automatically zero. In other words, the strand model eliminates all problems of QED; in fact, QED appears as an approximation of the strand model for negligible strand diameter. In passing, we thus predict that perturbation theory for QED is valid and *converges* if the strand model, and in particular the finite strand diameter, is taken into account. (The diameter is the only gravitational influence predicted to affect QED.)

The strand model also suggests that the difference between renormalized and unrenormalized mass and charge is related to the difference between minimal and non-minimal crossing switch number, or equivalently, between tangle deformations with few and with many crossings, where strands are deformed on smaller distance scales. In other terms, unrenormalized quantities – the so-called *bare* 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.

The strand model also suggests a visualization for the cut-off used in QED. The cut-off is a characteristic energy or length used in intermediate calculations. In the strand model, the cut-off corresponds to the size of the image.

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 change or 'running' with Page 380 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 up to energies just below the Planck scale. Deviations between QED and the strand model are only expected near the Planck energy, when tangles of Planck diameter are pulled tight.

THE ANOMALOUS MAGNETIC MOMENT

The anomalous magnetic moment g of the electron and of the muon is given by the well-known expression

$$\frac{g}{2} = 1 + \frac{\alpha}{2\pi} - O(\alpha^2) , \qquad (153)$$

where $g/2$ is half the so-called g-factor, with a measured value of 1.00116(1), and α is the fine structure constant, with a measured value of 1/137.036(1). Julian Schwinger discovered this expression in 1948; the involved calculations that led Schwinger to this and similar results in quantum field theory earned him the 1965 Nobel Prize in Physics. The result is also inscribed on the memorial marker near his grave in Mount Auburn Cemetery. The strand model proposes an intuitive explanation for this result.

Generally speaking, the factor $g/2$ describes the ratio between the 'mechanical' or 'geometric' rotation frequency – the rotation of the particle *mass* that leads to spin – and 'magnetic' rotation frequency – the rotation of the particle *charge* that leads to the magnetic moment. More precisely, the definition of the g-factor of a particle with charge e and mass m is

$$\frac{g}{2} = \frac{\mu/e}{S/m} . \qquad (154)$$

Here, μ is the magnetic moment and S is the intrinsic angular momentum, or spin.

The *mechanical* or *geometric* rotation frequency is related to the ratio of the intrinsic angular momentum L and the mass m. Using the definitions from classical physics, we have $\boldsymbol{S}/m = \boldsymbol{r} \times \boldsymbol{v}$. The *magnetic* rotation frequency is related to the ratio of the magnetic moment $\boldsymbol{\mu}$ and the electric charge e. Classically, this ratio is $\boldsymbol{\mu}/e = \boldsymbol{r} \times \boldsymbol{v}$. Therefore, in classical physics – and also in the first order of the Pauli–Dirac description of the electron – the two rotation frequencies coincide, and the factor $g/2$ is thus equal to 1. However, as mentioned, both experiment and QED show a slight deviation of $g/2$ from unity, called the *anomalous* magnetic moment.

In the strand model, the geometric or mechanical rotation of a charged elementary particle is due to the rotation of the tangle core as a rigid whole, whereas the magnetic Page 223 rotation also includes phase changes due to the *deformations of the tangle core*. In particular, the magnetic rotation of a charged elementary particle includes phase changes

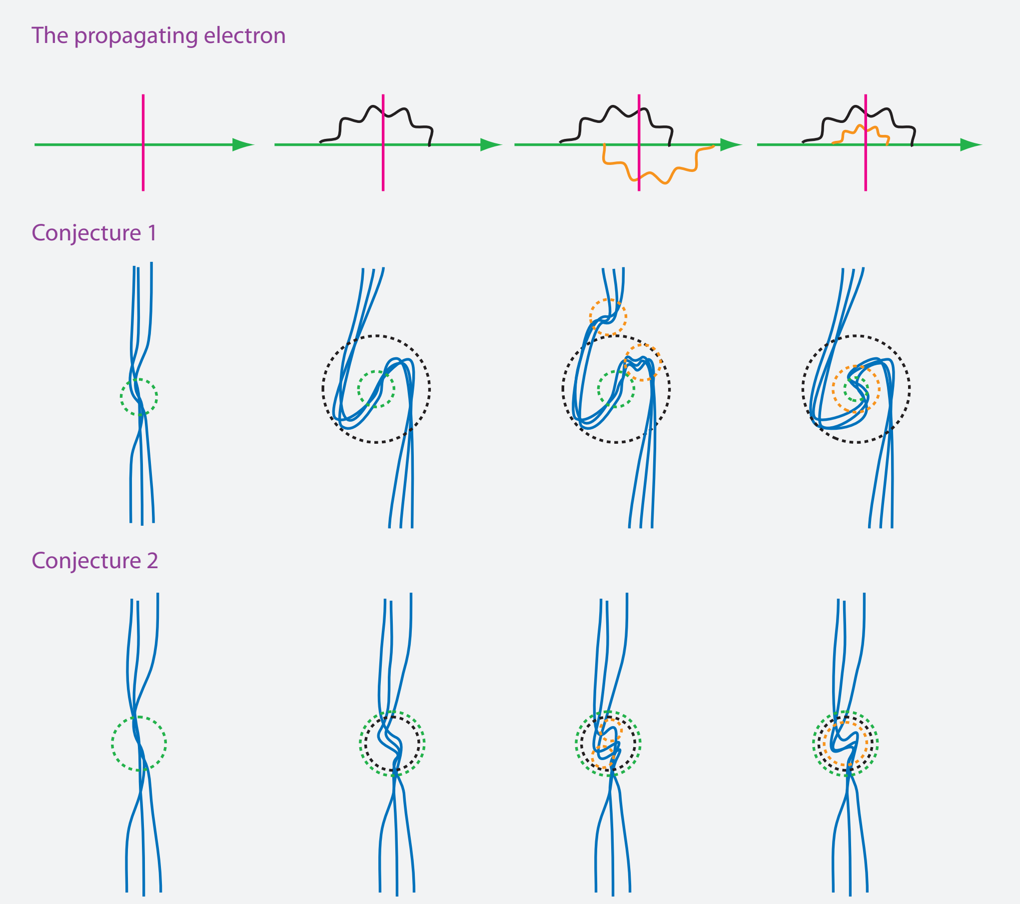


FIGURE 59 Two conjectured correspondences between the Feynman diagrams of quantum electrodynamics and the strand model for a propagating free electron. The lower strand model configurations are shown for a single instant – marked in magenta – of the electron propagator drawn above them. (For simplicity, the external field is not drawn.) In the first conjecture, the loops of the belt trick are conjectured to correspond to the virtual photons in the propagator and to be responsible for the anomalous magnetic moment. In the second conjecture, the deformations of the core correspond to the virtual photons.

due to emission and reabsorption of virtual photons, i.e., of twisted loops.

In nature, the probability of the emission and reabsorption of a photon is determined by the fine structure constant α. The emission and reabsorption process leads to an additional angle that makes the 'magnetic' rotation angle differ from the 'mechanical' rotation angle. Since the fine structure constant describes the rotation of the phase due to virtual photon exchange, the emission and reabsorption of a virtual photon leads to an angle difference, and this angle difference is given by the fine structure constant itself. The ratio between the purely mechanical or geometric and the full magnetic rotation frequency is therefore not one, but increased by the ratio between the additional angle α and 2π. This is Schwinger's formula.

In short, the strand model reproduces Schwinger's celebrated formula for the anomalous magnetic moment almost from thin air. The strand model also implies that Schwinger's formula is valid for *all* charged elementary particles, independently of their mass; this is indeed observed. Higher order corrections also appear naturally in the strand model. Finally, the strand model implies that the complete expression, with all orders included, *converges*, because the full result is due to the shape and dynamics of the tangle core. The discussions about the existence of the perturbation limit in QED are thus laid to rest.

If we look into the details, it might be that the belt trick itself is at the origin of the anomalous magnetic moment. A conjecture for this connection is proposed and illustrated in Figure 59: if the two loops formed by the belt trick are seen as virtual photons, the factor $2\alpha/4\pi$ arises naturally. So do the higher-order terms. This explanation would relate the belt trick directly to the additional magnetic rotation angle. However, it might also be that this correspondence of the strand images in the figure to the upper diagrams is not fully correct. The topic is subject of research. Challenge 156 e

A second conjecture is also given in Figure 59. The virtual photons could correspond to deformations of the tangle core. This conjecture is more in line with the distinction between gravity and gauge interactions given above, where it was stated that gravity is due to tail deformations and gauge interactions are due to core deformations. This conjecture is more in line with the distinction between a geometric and a magnetic rotation: the geometric rotation would be due to the rigid rotation of the tangle core, and the magnetic rotation would be due to an additional effect due to core deformation. Page 223

Both conjectures on the origin of the g-factor imply that $1 < g/2 < 2$; in fact, we can even argue, using $\alpha < 1$, that the strand model implies

$$1 < g/2 < 1 + \frac{1}{2\pi} \ . \qquad (155)$$

This is not a new result; it is already implied by ordinary quantum field theory. However, the strand description of particle rotation suggests a way to calculate the g-factor and the fine structure constant. We will explore this below. Page 387

MAXWELL'S EQUATIONS

The strand model also allows us to check Maxwell's field equations of classical electrodynamics directly. The equations are:

$$\begin{aligned}\nabla\, \mathbf{E} &= \frac{\rho}{\varepsilon_0} \ ,\\ \nabla\, \mathbf{B} &= 0 \ ,\\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \ ,\\ \nabla \times \mathbf{B} &= \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} + \mu_0 \mathbf{J} \ . \end{aligned} \qquad (156)$$

▪ The first of these equations is satisfied whatever the precise mechanism at the basis of twisted 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 $1/r^2$. This relation is valid independently of the underlying motion of the twisted loops, because space has three dimensions: 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 twisted loop emission rate for an elementary charge.

- The second of the field equations (156) 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 52. 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 (156): 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 52 – 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