The Universe of Codes

Beyond General Relativity and Quantum Mechanics

Andrea Gregori*

Information is what tells us about the world. Could we think of physical reality itself as consisting just of information? Could the universe represent over the time the most general collection of logical codes? If we think of information as expressed in terms of binary codes, we could interpret binary sets as spatial arrangements of energy, giving rise to discrete geometries. Can we then identify a path through geometries, or sets of geometries, representing the history of the physical universe? This book introduces a theoretical framework that, by giving an answer to these questions, unifies general relativity and quantum mechanics at a more fundamental level. The geometry of the universe, its relativistic and quantum mechanical nature and the spectrum of elementary particles show up as the consequence of an entropic principle. All the masses and interaction couplings are computed as functions of the only free parameter, the age of the universe. This approach contributes to shed new light on several branches of physics, from elementary particle physics to the physics of high temperature superconductors, to cosmology, as well as on certain aspects of the natural evolution.

This works updates all the previous research regarding the project „the universe of codes“ with a refreshed and improved presentation of the arguments, and refined calculations. It is addressed both to people new to these topics, and to those who already had a look at some of the previous works. People interested in an updated version of any of the works of this project are encouraged to have a look at the corresponding topics in this book. This publication aims also at providing an as much as possible self-contained discussion of the whole project. This should save the reader the trouble of jumping forth and back through references.

The content here presented is the **2022 update** of the book „The Universe of Codes“ (Lambert Academic Publishing) first posted here in 2019. Section 4.4 has been reworked, corrected from misprints and updated in order to incorporate more recent research results.

* www.universeofcodes.eu

The Universe of Codes

Beyond General Relativity and Quantum Mechanics

Andrea Gregori

2019 (v2022)

Contents

Preface

What is it about

This book discusses a new approach to the physics of space and time, which enables a unified description of elementary particles and gravity. Relativity, quantum mechanics, field and string theory are lifted to a new level of interpretation, that puts everything into a different perspective, starting from the very definition of space, and time. The most interesting aspect of this scenario is its predictive power. The type of elementary particles, the geometry of the universe, and its relativistic and quantum mechanical nature, are predicted as the consequence of a very elementary and basic entropic principle. All the masses and interaction couplings are computed as functions of the only free parameter, the age of the universe.

On a formal level, much is similar to what is already known, being firmly anchored in the theory of relativity, quantum mechanics, field and string theory. However, everything is incorporated into a theoretical framework rooted on more fundamental principles and assumptions than quantum mechanics and relativity. Not only all what already worked within these theories is here recovered, although just as an effective approximation, a special case valid under appropriate conditions, but it comes together with new aspects, that put under a different light also the issues related to the so-called "new physics", namely the theoretical explanation of new phenomena, detected in particle colliders or showing up in astronomical observations.

Despite the fact that many section titles, and names of physical quantities, evoke known concepts, there are here only few already known expressions, and even in case some may look familiar, as a matter of fact reading them extrapolated from their context can be misleading, because here everything is getting a new interpretation. Therefore, it cannot simply be extracted and directly compared with similar-sounding concepts of the literature. The best way to approach this work is to read it from the beginning, getting gradually the more and more familiar with the concepts which are introduced.

The background

This work collects, updates and improves, the results of several years of research. The project started with an investigation of non-perturbative string-string dualities, and ended up with the proposal of a new theoretical scenario, which entails a paradigm shift with respect to the usual approach to elementary particle physics, gravity, cosmology, with repercussions also in other domains of physics or natural science, such as high-temperature superconductivity, palaeontology, and potentially other ones. The first steps of the research have been set in [1], a work that extended some ideas previously worked out in [2] to a lower supersymmetric case of string string dualities, allowing to set the ground for a thorough investigation of non-perturbative aspects in a pure stringy context, i.e. not just relying on the extension of properties belonging to field theory. This was followed by [3] to complete the pattern of dualities by including also type I string. The basis for a thorough investigation were set in [4], a classification work of N=2 type II orbifolds. After [5] opened the way to the identification of non-perturbatively realized gauge groups, [6] provided a complete duality pattern within N=2 orbifold constructions. The turning point in the approach to the string target space was provided by the investigation of the cosmological constant, [7]. The ideas thereof matured, together with the non-perturbative pattern enlightened by [6] led to a first attempt to produce an organised picture in [8]. Many ideas subsequently developed were already there, although often confused, and sometimes not so consistently worked out. The following years have been therefore devoted to a deep reconsideration of the whole scenario, which led, step by step, to the point of view here presented. These steps include the investigation of the consequences of the time dependence of masses and couplings [9], and the recognition that the scenario that was being developed could be considered a generalization of the Feynman path integral to include the geometry of space [10]. The investigation of the long-time effects of time-depending energy scales on the interpretation of palaeontological observations [11] went in parallel with the investigation of the theoretical bases of the scenario from a more abstract point of view. This made clear that, step by step, things were evolving toward the construction of a new theory, underlying quantum mechanics and general relativity [12]. This had striking implications, ranging from cosmology (e.g. [13]) up to solid state physics, such as a new approach to high-temperature superconductivity [14]. An attempt to collect all this in an organised fashion was provided by [15], followed by [16] and [17], whereas subsequent research, including also [18], required an update, presented in [19, 20, 21].

After some years, during which, together with updates of previous works, also [22] was produced, the state of the art of this research is now presented in this more extended work, which is aimed at saving to the reader the time of going through references, and trying to collect pieces of information spread through the various updates of an in-progress research. This should also allow to get more clearly the unity and consistence of a research, in which all the aspects are parts of a unique picture.

Acknowledgements

H. C. Hege, D. Franco and G. Zuccaro-Labellarte deserve my warmest thanks for their friendly and invaluable advice on how to improve the readability of this work. This was particularly helpful in the introductory section.

1 Introduction

The search for a unified description of quantum mechanics and general relativity, within a theory that possibly describes also the evolution of the universe, is one of the long debated issues and open problems of modern theoretical physics. The research is focused on finding the solution of the problem, intended as an appropriate construction within a specific theoretical framework (field theory, string theory, quantum mechanics, or else). For instance, within string theory this means finding the right geometry on which to compactify the string in order to produce a spectrum of particles, fields and interactions that reduces to the one of the Weinberg and Salam Standard Model of elementary particles at the electroweak scale. This is a way of proceeding somehow by "trial and error", in the hope that, once the result will be obtained, it will also shed light onto unknown theoretical aspects and lead to a deeper understanding of fundamental physics. However, the fact that such a solution has not yet been found could be the signal that this is a wrong way of proceeding. Perhaps, in order to progress, one has first to go through more fundamental questions, and a path toward the solution can be found not by looking for a result, but for a set of fundamental requirements. Are all the conditions we usually associate with quantum mechanics and relativity really fundamental? Perhaps some of them are just effective approximations, valid in a specific range of parameters, of something more general and fundamental. Let us suppose that, by proceeding by trial and error, we find the right construction, the right model that gives us the physics we want. In that case, we can try to understand what does it make of it something special, besides the fact that it works. On the other hand, since, as a matter of fact, we are not able to find the solution, maybe it is wise to revert the argument, and try first to answer to the question: what would allow to select it "a priori", even before checking that it works? If we can answer to such a question, we may hope to get a hint on where to look for the solution. From this point of view, it may be not necessary, even misleading, sticking on the requirement of implementing in a unified theory the rules of canonical quantization on one side, and imposing the bound on the speed of light, and the Lorentz group of coordinate transformation, on the other side: both quantum mechanics and relativity could show up as a consequence of more general requirements and assumptions, and their "unification" could occur at a more fundamental level, namely at a level in which quantum mechanics and relativity are not yet "disentangled" from a description in which both are in principle contained. Their unification would then be based on the formalism of neither one of these theories. A hint that this could be the case is provided by the fact that general relativity, and therefore gravity, cannot be consistently quantized, in the sense of applying the rules of quantum mechanics to some kind of field theoretical description of the gravity zero modes, without encountering troubles. The relation of this hypothetical underlying theory to quantum field theory and general relativity could be similar to the transition from classical to quantum mechanics. There, the classical description is only recovered as a large scale approximation, because on the fine detail there is indeed a gap between the two descriptions.

String theory is widely considered to be the most legitimate candidate for building up a "theory of everything". However, as it is it is an "empty" theory: there is no rationale for a choice of the geometry for its coordinates, other than the simple looking for a compactification that produces the right spectrum (an "a posteriori" justification). By the way, such a compactification has not been identified. This may seem a rather technical issue, a matter of just further looking for. However, the failure in finding the right geometry could hide a deeper problem, namely the fact that our expectations about what we must find are wrong. This is a rather subtle point, so I try to explain it more clearly. The common lore is that one must eventually find the spectrum and the interactions of the Standard Model of elementary particles, a well tested and good working field theoretical model. Indeed, the Standard Model works well in a certain subset of cases, but presents some rather critical points and its predictive power is rather limited. Since all its parameters are basically free, in most cases a test of the model verifies the capability of fitting a set of experiments within an appropriate choice of values of its parameters. This is certainly a non trivial fact, nonetheless theoretically unsatisfactory, because it is not a test of its pure predictive power. However, if it is true that in a rough approximation the degrees of freedom of the Standard Model can be safely considered as the basic ingredients of elementary particle physics, it is not necessarily true that so must be considered also the *details* of their interactions. In particular, if the string scenario is not simply an extension of a field-theoretical scenario, it is not obvious that it must reproduce all the fields which are required in order to make the field theoretical description of elementary particles, and their interactions, consistent. Indeed, the field theoretical description not only fails in providing a unified theory producing correct experimental predictions (not just partial data fittings), but seems to show also some inconsistencies. In the purpose of looking for the basic principles on which to base an underlying theory, a look at these is a step that can not be avoided.

Some critical points

To start with, one may wonder whether the entire description of **masses** is plagued by a fundamental inconsistency: masses are source of gravitation, therefore there cannot be a consistent (i.e. self- contained) quantum theoretical description of massive states without a quantum theory of gravity (of course, beyond this first-order evidence, one can also point out that already the fact in itself of speaking of energy of these degrees of freedom implies by consistency also speaking of gravity). Another possible theoretical inconsistency is given by the assumption of working in an **infinitely extended space(-time)**. When combining the principle of causality, the finiteness of the speed of light and the existence of a bing bang, i.e. a temporal origin of the universe, it seems that requiring the physical description at any finite time to be embedded in a space of infinite extension is in some sense redundant, requiring more than what we need. How can one test physics beyond the horizon set at a distance from the observer corresponding to the distance travelled by light since the origin of the universe? How can things that exist outside the **causal region** of an observer influence/affect the physics he observes, his experimental observations/detections? Does it make sense to require such an infinity condition? Of course, "beyond the causal horizon" phenomena can be precisely the domain of a quantum theory. But then, on which basis are these extra regions parametrized in the same way as the space-time within our causal region, namely, in terms of classical coordinates to be treated as parameters of an action, even in the case of a quantum field-theoretical framework? On the other hand, abandoning this apparently redundant, unnecessary requirement eliminates many regularization problems, and puts in a completely different light several issues, from the type of symmetries the fundamental physics possesses (translation, time evolution), up to the fundamental description in terms of fields, together with the entire formalism they imply (infinity/zero momentum regularization) etc...From absolute principles they get immediately downgraded to mere approximations.

Since all these considerations are a consequence of *classical* arguments (classical geometry, classical concept of causality, relativity), quantum mechanics can in principle break this chain of implications. It could therefore be that what has to be considered of finite extension is just the classical part of space, provided one understands/properly defines what "**classical**" and "**quantum geometry**" means. If quantum geometry is intended as the quantum fluctuations produced by the propagation of a canonically quantized classical field, such as the graviton, a quantum universe, and a quantum space, cannot be much different from its classical part: the quantum fluctuations of the geometry are in any case based on a classical concept of space. However, if one thinks of the classical space as the one whose horizon is "stirred" by the propagation of massless fields (again, the graviton, or the photon, or alike), and thinks of quantum mechanics somehow in the light of the Feynman path integral, as the sum over all paths, including tachyonic ones, it does no more appear unreasonable that to a classical space of finite size does correspond a quantum counterpart of infinite extension. The quantum geometry, whatever its precise definition may be, is in this case not just a perturbation of a classical geometry: it may be something that destroys the classical idea of space-time, and the very idea of coordinates.

On the other hand, as it is defined, the Feynman path integral does not seem suitable for a quantum theory including gravity, i.e. the geometry itself of space-time. Being defined as a sum weighted by the action, everything results to be triggered by an **action principle**. This could be fine (and it is) as long as one does not bring into play also the space itself in which the fields concurring to build up, with their dynamics, the action, do live and propagate. An action is a function of fields whose parameters are the coordinates of the space they live in. As such, the action is indirectly a function of the geometry. However, summing up over all geometries weighted by an action produces a staple which corresponds to a certain energy. As such, it implies a modified geometry. The problem is therefore non linear, and in principle non-perturbative. Moreover, one may wonder if thinking of quantum gravity more or less as of a theory of quantized fields like the graviton catches the deep essence of a quantum space-time. If one thinks in terms of a space-time whose coordinates are themselves quantized (and/or promoted to quantum fields), already the idea of basing the physical description on the concept of action becomes in itself meaningless.

A turning point

Should we then abandon the idea of action as the basic principle on which to build any dynamics of space-time? Is time really to be considered on a similar footing of space, as special relativity seems to suggest, or is the construction of a space-time just an elegant but approximated way of dealing with two things (space, and time) which are however conceptually very different? In other words, is the flip in the metric signature from Euclidean to Minkowskian the result of a somehow “dynamic” symmetry breaking, or does it signal a deep conceptual difference?

Let us try to go back some steps, and see where is it possible to find solid land. From general relativity we learn that a geometry is in itself equivalent to a distribution of energy. Let us stick on this point, which looks very solid, and work it out more deeply. According to Einstein’s relativity, also masses are energy packets. **Everything** in the world can therefore be viewed **in terms of energy**. Dynamics, and interactions, are then equivalent to changes of the local, and global, geometry, occurring during time. At any instant of time, the universe itself can therefore be viewed as a static point-wise assignment of energy amounts along space. Quantum fluctuations and uncertainties can then be viewed as fluctuations among such a kind of static assignments; dynamics, and time evolution, as a progress through static energy assignments. Let us call each set of such point-wise assignments, for each point of an infinitely extended space (which of course for the moment is not ~~yet~~ a geometric space), a “geometry”. In principle, each geometry potentially describes a universe. A geometry does not have dynamics in itself (it is just a static assignment), does not contain fields and/or particles or whatever degrees of freedom. However, a **history of geometries** can in principle correspond to a physics of evolving degrees of freedom such as particles and/or fields. The only intrinsic property a static geometry can have is the recipe of its assignment of energy along space, to which one can associate an **entropy** in a natural way: if we think of a universe ruled by a principle that just blindly throws energy units on a target space, the most entropic geometries will be those that possess a higher amount of symmetry. Of course, before being properly defined, all these concepts are just loose words, but these loose ideas served as starting point and reference guide for developing the research that we are going to present and discuss in this work. They can be viewed as an embryonic form of change of paradigm: instead of looking for a geometry of space in which to frame a set of fields, we try to define a universe in terms of spaces intrinsically defined by geometries. It is like to start from a set of static pictures, and see if, and in which sense, we can get a story by grouping and interpreting them as the frames of a movie.

A history of geometries

Let us then make the hypothesis that the universe is the set (the collection) of all geometries, intended as the collection of all the random assignments of a certain amount of energy to a certain space, simply ruled by the fact of being assigned randomly. We don't have therefore a rationale, except from a very basic and elementary, I would say generic, one: there is no recipe, no rule, other than just the idea that "all what can exist exists". In this "whole" we don't impose a selection mechanism: we want rather to see whether in the "chaos" of "everything is there" there are structures, produced by the simple fact that, perhaps, certain geometries, or parts of geometries, are statistically favoured, i.e. to see whether, in the "blind throwing energy all around", some configurations occur more often than others. If this occurs, by a simple statistical averaging these structures will determine the dominant shape, and perhaps also the space dimension, of the universe. To this regard, it is legitimate to ask whether the simple blind, random throwing of energy just ends up with a space evenly covered by energy. Naively, one would say this is precisely the result: after all, it is like saying that tossing a coin produces, on a large number of tosses, simply 50% heads and 50% tails. However, this occurs because we have a recipe to distinguish not just tail from head, but also top from bottom. Therefore we can say: "the result is what comes out on the top", and this can be either tail or head. But imagine we do not have a way of distinguishing top from bottom... Indeed, top can be distinguished from bottom only if there are asymmetries in the universe, i.e. points with respect to which top and bottom are not identical. In the absolutely empty space, there are no such asymmetries, and top is not distinguishable from bottom. In our problem, the target space does not have in itself reference points allowing to distinguish between two energy distributions which are just "displaced" with respect to each other. Energy distributions are classified through their only built-in property: their internal degree of symmetry. Reference points are defined by built-in asymmetries of a geometry, and introduced in the space together with asymmetric energy assignments. Let us now fix the amount of energy we throw on the target, and think of stapling all the geometries on top of each other, starting from the most symmetric one. Proceeding in the stapling by steps of decreasing symmetry, by using the asymmetries introduced by the previous geometry as reference point for placing the next geometry one builds up a space in which all symmetries are broken. The space obtained by superposing in this way the various geometries will have a symmetry not larger than the intersection of all the symmetries of the stapled geometries. Each further step in the stapling increases the breaking of symmetry [1]. Figures 1.1 and 1.2 illustrate this principle with an example. Of course, all this has to be set more precisely, and it does make only sense after we have specified some more details. For the moment, just accept the idea, to see where we are going to. Indeed, what we are doing is trying to see whether there can be a "built-in" way of weighting geometries, which does not require the introduction of external inputs such as an action, maybe together with further requirements.

In this scenario, the existence of the universe is itself part of the game: if we include in the possibilities also that of not throwing energy at all, this is just one of the many logical possibilities, clearly statistically much less relevant than the existence of a geometry, just because there are many more logical possibilities of arranging a non-vanishing amount of energy than of having no energy, and therefore no geometry, at all. For the same reason, higher energies will produce more geometries than lower ones. In particular, subregions of total energy E' of a geometry corresponding to the distribution of a total amount

[1] Here is the key difference as compared to examples like the coin toss: in that case, the coin is considered as inserted in a space working as external reference frame, here we are defining and building the space, and the reference frame as built-in into the space, through intrinsic properties. It is like saying that top and bottom are defined by the coin itself, which knows only about "head" and "tail". Top and bottom can then only be defined in terms of head and tail. If we call "top" the head, the head will all the time staple to head, to a 100% of head results, by definition, and we will never end up with a 50/50 averaged heads-tails result.

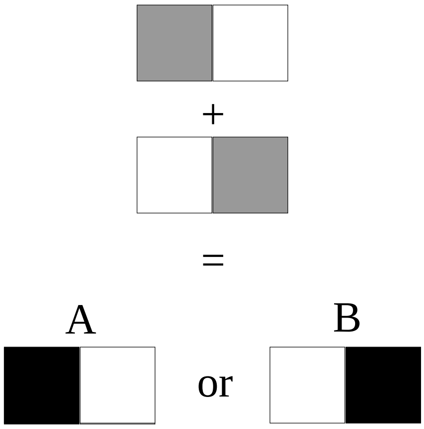


Figure 1.1: The superposition of two geometries produced by assigning one energy unit to a space consisting of two units. A and B are here *the same* geometry, because there is no external point to serve as reference for a rotation, thereby enabling to distinguish A from B.

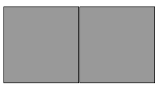


Figure 1.2: This is not the superposition of two geometries as in figure 1.1, but one geometry given by the assignment of two energy units to a space consisting of two space units.

of energy E (where, obviously, $E \geq E'$) can be viewed as geometries corresponding to a total amount of energy E'. In this sense, we can say that the set of all the geometries corresponding to a total energy E "contains" the set of all the geometries corresponding to a total energy E'. This allows to associate an ordering through sets of geometries. Indeed, one is tempted to call each set of all the geometries at a given total energy E a "universe" $U(E)$, and view the path through increasing total energy as a time progress, a history of the universe, through the identification of the total energy with the time: $U(E) \to U(t)$. We obtain in this way a time ordering. The zero energy geometry, the "non-existence of the universe", is then just the trivial starting point. It does not make sense to ask "how long" the universe "did not exist" or, "pre-existed", because, seen in this way, time is a built-in property of the existing universe. So, we have now a "history" out of static frames. However, the movie we obtain is not of the classical type: instead of having a sequence of frames, we have a sequence of collections, or staples, of frames. Can this be somehow considered a "quantum movie"? Let us go on with the investigation of the universe at each finete time/total amount of energy.

The Planck length

In order to investigate a possible physical content, we must first of all introduce units of measure. For instance, by introducing a unit for the energy, and then conversion constants to convert to time, space lengths, etc... As it was for the case illustrated by the coin, here too we must pay attention to not fall into the mistake of considering the space, and its content, as immersed into something else, implicitly assumed to play the role of reference system. This would produce the misleading impression of being allowed to introduce a unit of measure without necessarily breaking scale invariance. Introducing a unit of scale *breaks* scale invariance, because by definition it introduces a reference point in a scale, therefore a preferred size: as we have seen when talking about geometries, reference points are intrinsically introduced only by the breaking of symmetries. It is easy to realize that the only way of introducing a unit of length, or energy, in an absolute, intrinsic way, is to introduce a *minimal* length and a minimal energy. These are such that any length, and any energy amount, can only be an integer multiple of the unit. Measuring is therefore not just "comparing to", but "counting". Otherwise, we would have a floating scale, completely meaningless until, through some mechanism introduced "ad hoc", one breaks scale invariance and sets a certain type of asymmetry as a reference point. This is what is done "a posteriori" in any physical theory build around certain experimental observations, in which some objects are taken as reference points. One works in a world in which symmetries are already broken, and tries to write a symmetric theory together with scale units that only make sense in a regime of broken symmetry. This is therefore a point that deeply distinguishes our approach, which instead aims at constructing a theory based on intrinsic, almost "self defining" properties, in the search for the minimal set of inputs one must impose to the most general set of axioms, in order to produce a meaningful physical world.

Introducing a minimal size allows to *count* energy, space lengths etc..., without the possibility of (making this operation trivial by simply) re-absorbing any rescaling into a redefinition of the units of measure. In this way we also obtain a true, meaningful *history*, or time progress. Indeed, this is equivalent to saying that we assume to work in a discrete world: space is discrete, with a lattice length unit, and energy is also discrete, with an energy unit. Intuitively, one would think that, since integers are a subset of the real numbers, we are in this way truncating the range of values, and the precision in our description of physics. The question is therefore: does this really constitute a limitation in the investigation's power of a theory? Is the range of the physical phenomena reduced? A closer look at number theory tells us that, as a matter of fact, real numbers are *constructed* through a chain of arguments (limit procedures, Dedekind sections, etc.) which are all based on the natural numbers. Although it may sound strange, from a logical point of view this means that the information content of real numbers is not larger than that of natural numbers. The key point is the extension of the theory to include the infinite. Starting from a theory defined on the discrete, but allowing the range of values to run up to infinity, we effectively recover many aspects of a theory described on the continuum. The key difference is however that, in this case, the continuum is viewed as an approximation of the discrete, and not the other way around. In particular, a whole bunch of technical problems of field theory, including ultraviolet regularization, are absent by construction. But also the fact of working with a universe characterized by a finite volume geometry at any point of its history implies looking at several aspects of field theory from a completely different perspective (see e.g. the issues related to topological monopoles, infrared regularization, etc...). Up to conversions, this unit can be used to *define* the Planck mass.

Once introduced a unit, we can state that any geometry corresponds to the distribution of a finite amount of energy. It will therefore describe a universe of finite extension. Nonetheless, since at any "time" the number of such geometries is infinite, there is no bound neither to the extension of space, nor to its dimension.

Classic and quantum space

Working with discrete and finite quantities allows to compare symmetries, and volumes of symmetry groups, and therefore define the stapling of geometries, unambiguously, by introducing in a natural way a weighted superposition of geometries. The weight of each geometry, i.e. its weight in the phase space of all the geometries, is naturally chosen to be proportional to the volume of its symmetry group. One can therefore also speak of entropy of a geometry, introduced as usual as the logarithm of the weight. Once all this is properly set up, geometries can be compared through their entropy, and we get an insight into the statistically averaged shape of the universe at each step of its evolution.

In this approach, there is in principle no distinction between "geometry" and "objects living in a space with a certain geometry": everything is a geometry, and it belongs to our interpretation to single out aspects that we assign to what we call "background space" and what we call "matter or field degrees of freedom" that move and interact within, and with, this space. It turns out that the most entropic, and therefore dominant, space dimension is three, with dominant geometry of the universe the (discrete approximation of the) 3-sphere. This is in fact the most symmetric geometry, and the one with the highest entropy/energy ratio at fixed radius among the spheres of any dimension. It turns also out that the less entropic geometries weight much less. These facts allow us to assume the 3-sphere to represent the classical geometry of the universe, on top of which the tower of less entropic, less symmetric geometries, contribute to build up a shaped world in which all symmetries are broken. The dynamics will result from the tracking of the modifications of these shapes through the history of the universe. The infinite staple of the more and more singular geometries implies the existence of shapes which are quite far away from what can be interpreted in terms of classical objects, and their motion. Indeed, the whole staple accounts also for what we will identify as quantum mechanical fluctuations of the classical space.

This construction seems therefore to implement a solution to the remarks we raised at the beginning, namely having a classical space bounded by a causal horizon, while allowing a quantum space to extend out of the causal region.

The speed of light, the Heisenberg's uncertainties

With the previous definitions, the "universe" is set up: we don't need anything else than looking at what are the implications of the founding statements. These imply that the entire information about the universe is encoded in a partition function given by the sum over all geometries ψ, weighted by their entropy S:

$$\mathcal{Z} \;=\; \sum_{\psi} \mathrm{e}^{S(\psi)} \,. \tag{1.0.1}$$

Notice the "measure" of this sum: at the exponent there is not an action, but trivially the entropy, defined as the logarithm of the weight, the statistical weight, corresponding to the volume of the symmetry group of a geometry, intended as an assignment of energy units with certain symmetry properties. The sum 1.0.1 can be rearranged as:

$$\mathcal{Z} = \sum_E \mathcal{Z}(E)\,, \tag{1.0.2}$$

where

$$\mathcal{Z}(E) = \sum_{\Psi(E)} \mathrm{e}^{S(\Psi(E))}\,. \tag{1.0.3}$$

In this way, the "time" dependence is made explicit. Expression 1.0.1 therefore does not look like (an extension of) a Feynman path integral. Nonetheless, their equivalence can be shown to hold in a "classical geometry" limit.

The partition function 1.0.1 implies that all what we observe is given by a superposition of geometries, and whatever value of observable quantity we can measure is smeared, is given with a certain fuzziness. Evaluating the contribution of non-maximal entropy geometries at any total energy step allows to see that the smearing quantitatively corresponds to the Heisenberg's inequality. This suggests the possibility of interpreting this scenario in terms of quantum physics. Indeed, an inspection of the geometries that arise in this scenario, the way "energy clusters" arise, and their possible interpretation in terms of matter, particles etc., allows to conclude that 1.0.1 can indeed be viewed as formally implying a quantum scenario, once the Heisenberg's uncertainty is given a *new interpretation.* The Heisenberg's uncertainty relation arises here as a way of accounting not simply for our ignorance about the observables, but for being these quantities ill-defined in themselves: all the observables that we may refer to a three-dimensional world, together with the three-dimensional space itself, exist only as "large scale" effects. Beyond a certain degree of accuracy they can neither be measured nor be defined. The space itself, with a well defined dimension and geometry, cannot be defined beyond a certain degree of accuracy either. This is due to the fact that the universe is not just given by one geometry, the dominant one, but by the superposition of all possible geometries, an infinite number, among which many (an infinite number too) don't even correspond to a three dimensional geometry. This interpretation of quantum mechanics does not contradict any of the conclusions of the traditional interpretation: from a practical point of view, on the physical systems in which traditional quantum mechanics is known to work, they lead to the same results. However, this new approach works also in more general cases, and potentially includes not only cosmology and gravitation, but the basic structure and concept of geometry of the universe.

But there is more to the universe defined by 1.0.1. Inspecting the rate of propagation of maximal entropy paths inside the universe of geometries, as compared to the rate of energy/time evolution, allows to identify the equivalent of the relativistic bound on the speed of light. This is the maximal speed of propagation of <u>coherent</u>, i.e. non-dispersive, information (tachyonic configurations also exist and contribute to 1.0.2: their contribution is collected under the Heisenberg's uncertainty). The existence of this bound together with the Heisenberg's uncertainty, allows us to recognize that this scenario incorporates both relativity and quantum mechanics.

The link between a basically classical theory of geometries (although defined on the discrete) and the interpretation in terms of quantum mechanics is here provided by the inclusion, and stapling, of an infinite number of geometries at any step of the evolution of the universe. The consequence is a new kind of dynamics, of which the classical, and the probabilistic one of quantum mechanics, turn out to be approximations valid in appropriate limits.

The dynamics

The dynamics implied by 1.0.2 is neither deterministic in the ordinary sense of causal evolution, nor probabilistic. We may rather call it "determined", although impossible to know beyond a certain degree of approximation. According to our definition of time and time ordering, at any time the actual superposition of geometries does not depend on the superposition at a previous time, because the actual and the previous shape of the universe trivially are the superposition of all the possible geometries at their time. Nevertheless, on the large scale the flow of the mean values of observables through the time can be approximated by a smooth evolution that we can, up to a certain extent, parametrize in terms of the familiar concepts of motion and time evolution. Since the stapled geometries are weighted by their entropy, the evolution is driven by an entropic principle. As it is not possible to exactly perform the sum of infinite terms of 1.0.2, and it does not even make sense, because an infinite number of less entropic geometries don't even correspond to a description of the world in terms of three dimensions, it turns out to be convenient to accept for practical purposes a certain amount of unpredictability, introduce probability amplitudes and work in terms of the rules of quantum mechanics. These appear to be precisely tuned in order to embed the uncertainty, that we formally identified with the Heisenberg's uncertainty, into a viable framework, which allows some control of the unknown by endowing the uncertainty with a probabilistic interpretation. From this point of view, we can therefore give an argument for the necessity of a quantum description of the world: quantization appears to be a useful way of parametrizing the fact of being the observed reality a superposition of an infinite number of "states". Once endowed with this interpretation, this scenario provides us with a theoretical framework that unifies quantum mechanics and relativity in a description that, basically, is neither of them: in this perspective, they turn out to be only approximations, valid in a certain limit, of a more comprehensive formulation.

Spectrum and masses

In order to extract the physical content of 1.0.2 in terms of the familiar concepts of fields and particles, it is convenient to map into into a string scenario, something possible in the continuum limit, once the tower of geometries has been interpreted in terms of quantum fluctuations around a reference geometry. Under these conditions, string theory proves to be a legitimate approximation . However, the string construction itself is no more "free", as complained at the beginning of this discussion: it must be consistent with the theoretical premises of this scenario, of which it must constitute a representation. Therefore, in analogy with 1.0.2, instead of looking at just one string model, we will consider stapling any possible string compactification. The properties of the physical spectrum will not be determined by just one string realization, to be considered as "the right" string compactification, but will come out as the average result of an entropy-weighted sum over all string compactifications.

The usual approach to string theory imposes the compactification of a certain amount of "internal" coordinates in order to achieve a description with a four-dimensional infinitely extended target space, to be identified with the physical space-time. In the light of the previous discussion, if we want a correspondence with the universe of geometries we have no reason to require the space-time to be infinitely extended: also the four-dimensional space-time must be of finite extension, i.e. compact. This requirement puts the space-time on the same footing as the rest of the string target space even after string compactification. There is no external intervention to single it out. What is then the distinction between internal string space, and space-time, in an anyway compact space? We identify space-time as the part of string target space in which there are moduli, i.e. degrees of freedom corresponding to freely adjustable coordinates: they are the candidates to describe an expanding universe in a scenario in which the expansion is not driven by a built-in time-dependence of fields, but by an ordering like the entropy-driven total-energy ordering of the geometries we have just introduced. This is not the case of a twisted space. Stapling string compactifications, the dimension of space-time will therefore be set by the intersection of the "non-twisted" spaces, i.e. by the minimal surviving amount of coordinates whose moduli are not frozen. Is this a vanishing set? In order to answer to this question, it is not enough to look at the single string constructions: it could be that the coordinates we are looking for are simply non-perturbatively realized. However, if one admits that, through string-string duality, one can investigate the properties of the underlying theory of which the various string constructions represent dual slices, then such an analysis becomes possible, and it turns out to statistically favour precisely three space coordinates, suitable for building-up a four-dimensional space-time. This is a rather remarkable fact, which parallels the fact that, by stapling geometries, one does not obtain a shapeless, but a shaped world: a similar argument works not only for the set of geometries referred to a single space dimension, but also when comparing different space dimensions.

In such a world, supersymmetry turns out to be broken at a string-string duality invariant scale, to be identified with the Planck scale. Space is therefore curved, as it must be if it has to correspond to a universe which is basically a 3-sphere. The lack of symmetry under translation of the space-time coordinates due to the space-time compactness implies a different normalization and therefore also a different interpretation of string amplitudes. Now they are no more expected to compute terms of a Lagrangian density, but global quantities in the string space. In order to be compared with the usual terms they must be divided by a space two-volume. In this way, a non-supersymmetric vacuum with supersymmetry broken at the Planck scale is exactly what is needed in order to produce the correct value of the cosmological constant, which in this framework is not a constant, but depends on the age of the universe through its dependence on the classical volume of the universe. This energy density precisely corresponds to the amount of energy required by consistency in order to describe a sphere. On the other hand, since field theory is here not a consistent theory but just an approximation, supersymmetry is no more required in order to stabilize mass scales: the latter, as well as the entire spectrum of elementary particles and fields, are obtained through an analysis of the relative weights in the phase space of all the geometries by passing through the string representation of the set up.

The spectrum of particles and fields is obtained by investigating a set of string duals. Being forcedly, by construction, *perturbative*, they account for the physics on the *tangent* of the physical space. Owing to the flatness of the tangent space, supersymmetric string constructions come back into play as the string constructions in which to investigate the elementary excitations of matter, and of the massless fields (graviton, photon) that "stir" the horizon of space. Moreover, any perturbative construction corresponds to a decompactification limit, in which, at least in some directions, space is infinitely extended. For this analysis we can therefore use much of the known formalism and tools of string and field theory. This allows to obtain information about the degrees of freedom at a massless level. Masses are investigated at a different level, after re-introduction of the information of space compactness: they originate in fact from non-vanishing ground momenta of a compact space. A mass hierarchy arises from the fact that different particles originate from the breaking of symmetry produced by the stapling of string compactifications with a different weight in the string phase space: the same ground momentum contributes therefore to the mass with a different weight. The origin of mass scales is therefore non-field theoretical; their values must be considered as "on-shell", non renormalizing, and their stability is ensured by definition/construction. Through the dependence of the ratios of weights and of the ground momenta on the time-expanding classical size of space, masses turn out to depend on the age of the universe.

It is always possible to write an effective action for the interaction of particles and fields. However, this has nothing fundamental, it is just a convenient parametrization of a non field-theoretical scenario, valid around a certain point of the evolution of the universe. As such, it does not need to be self-contained and consistent. In particular, it does not need to be renormalizable as gauge theory with massive states. As a consequence, there is also no need of a Higgs mechanism, or of field-theoretical mechanisms that allow to dynamically stabilize symmetries and parametrize their breaking, such as for instance the Peccei-Quinn mechanism and Axions, etc...

If the field-theory effective action is not self-consistent, and string amplitudes undergo a deep re-interpretation; if, moreover, any string construction is just a perturbative slice of a basically non-perturbative theory, how are then computed scattering and decay amplitudes? At least conceptually, they must be computed by going back to the very basic definition of physical evolution in terms of staples of geometries evolving toward more and more entropic configurations. This means investigating the relative phase space weight of the corresponding phenomena, something rather non-trivial. Luckily, a lot of work can be recovered from traditional approaches, which remain a valid tool, even though they are no more the final truth. In practice, we will use a kind of mixed approach: starting from the known tools provided by an effective action, and/or the tool of loop string amplitudes, one must all the time keep in mind the limitations of these approaches, and be ready to deviate/correct/abandon them, as soon as they lead to contradictions with our premises, namely the entropic theoretical framework. At the time being, we do not have a general recipe: we will illustrate case by case how this can be successfully carried out.

Only one free parameter: the age of the universe

A point of strength of this scenario is its predictive power. There is no free parameter, everything being determined as a function of the only running quantity, the total energy, or equivalently the age of the universe. It results a theoretical framework that, in principle, can be used to predict the inputs of whatever effective theory, and determine their value. This includes the couplings of the elementary particles, like the electromagnetic, weak, and strong coupling. Since the effective dynamics is driven by an entropic principle, the type and strength of the couplings must be investigated by looking at the phase-space volume of interaction processes: the larger is the increase of entropy due to a certain interaction, the higher is its strength. The couplings originate therefore as ratios of volumes of a time-evolving set up, and, like masses, turn out to depend on the age of the universe. As such, they enter an effective action as parameters which must be "updated" as the universe evolves. In this way, an effective action always approximates the physics of one step of the evolution of the universe. As such, it may work well in describing "instantaneous" interactions, but cannot accurately account for phenomena that involve changes in space and time, such as for instance the breaking of the time-reversal symmetry. In this scenario there is by definition no symmetry under time reversal. However, an effective action, being built on a set of fundamental massless fields and particles of a time-invariant spectrum, plus mass and coupling terms, whose values reflect the situation at a certain point of the evolution of the universe, does not automatically contain all the appropriate symmetry-breaking terms needed in order to parametrize the time evolution. For instance, time-reversal or parity breaking must be introduced "ad hoc" by appropriate mechanisms, that just "mimic" the basic absence of these symmetries at the fundamental level, where the theory is not a field theory.

Comparing with experiments

Once the theoretical framework is set up, and we got the mass of all the elementary particles, and their coupling strength, expressed as functions of the age of the universe, we have a highly constrained set of predictions that we can test against the experimental results. To a first degree of approximation, the test is quite easy: it just requires to plug into the various expressions the value of the age of the universe converted in Planck units. The output is in astonishing agreement with the experiments. Owing to the absence of freely adjustable parameters, this is a rather strong result. A comparison up to arbitrarily higher order of precision is on the other hand more problematic, because in most (all) cases the experimental results are obtained through very refined data fitting procedures carried out within a well defined theoretical-phenomenological scheme. A true precision test of our predictions should be carried out in a similar self-contained way within our theoretical scheme, something only possible after equivalently refined computational tools have been developed also for this scenario, something that goes far beyond the scope of our preliminary investigation. So, for the time being, we must content ourselves with a mixed analysis, relying on the fact that, within our scenario, we are able to identify in first approximation the degrees of freedom for which the experimental results are provided. In this way, we are not only able to recover the fact that the value of the fine-structure constant is $\sim 1/137$, but also give an explanation of why this quantity is precisely of this size: the answer is that there is no deep reason to produce such a number instead of, say, 1/52, or 1/18700000...We measure this value because we measure it now, at a specific point of the evolution of the universe for which this is the value of the fine-structure constant, which is not a constant. Indeed, its value could be used as a measurement of the age of the universe. The same goes for all the other quantities, for instance the masses of the elementary particles. Indeed, in our investigation we first use the cosmologically derived value of the age of the universe in order to roughly compute the value of the neutron mass. However, since the neutron mass is given with a higher experimental precision than the age of the universe itself, we then revert the argument and we use the experimental value of the neutron mass in order to give a better estimate of the age of the universe, from which to compute all the other masses, and couplings.

Symmetry breaking, Higgs, Axions, dark matter

Owing to the stapling of geometries of any degree of symmetry, in this scenario all symmetries are broken. The massless fields of what in field theory are identified as unbroken symmetries, namely the graviton and the photon, are massless here too. They parametrize in fact the vectorial and spinorial part of the expansion of the geometry on the tangent, flat space associated to any point of the universe. Being at any time the classical volume of the universe finite, quantum fields must be considered as living in a box. Therefore, they always have a non-vanishing ground momentum. In the case of massless states, the ground energy/momentum scales with the inverse of the age (and therefore of the classical radius) of the universe. For the massive fields and particles it scales instead like some root of the inverse of the radius. Massless states can therefore extend as much as the entire classical universe (the largest classical extension possible in this scenario, what substitutes the infinite extension of field theory embedded in an infinitely extended space-time). Massive states are instead "localized" to a shorter scale.

We said that masses are not generated through the interaction with a Higgs field. One may object that the Higgs field has been experimentally detected. Therefore, although interesting from a pure theoretical point of view, this scenario is ruled out by experimental evidence. A further astonishing aspect of this scenario is that it does not predict the existence of dark matter either. If the goal of this approach is to reproduce the interpretation and parametrization of physics in terms of the Standard Model's degrees of freedom, or of extensions thereof, here we are definitely wrong. However, if the goal is to explain, and, perhaps, predict experimental observations, things change deeply. This scenario *predicts* the resonance around 125 GeV, and accounts also for the phenomena usually referred to the presence of dark matter. Simply, it explains them differently, through effects that do not exist in the traditional field theoretical approach. For instance, the resonance at 125 GeV (indeed, a bunch of resonances close to each other) is the effect of the existence, for any interaction, in particular in this case for the electromagnetic interaction, of S-dual components, or phases, due to the non exact breaking of T-(S-)duality (indeed, in this scenario no symmetry breaking is exact but only statistically realized). This allows to observe, at specific energy thresholds, scattering resonances due to the production of electromagnetically strongly coupled states. A similar phenomenon occurs also for the strong force: in this scenario, the main phase of the colour coupling is strong (coupling > 1), therefore pure strong coupling phase with no possibility at all of observing the elementary degrees of freedom, all confined into inseparable singlets. However, there is also a small amount of an S-dual phase, which accounts for the possibility of experimentally observing, at least in part, the internal structure of hadrons, as experiments indicate. The presence of two phases, with a lower weight of the "free field" phase, justifies also the lower importance of the colour interaction in the phenomenon of CP violation, making unnecessary explanations based on Axions fields. The aspects usually attributed to the presence of dark matter are even more exceptional, and related to the very particular quantum-relativistic geometry of the universe, and its interplay with our perception/detection of information, carried by photon and graviton.

Cosmology, and cosmological constraints

By definition, this scenario describes the physics of a universe centered around the observer: the extension of the classical space is a concept related to the observer. This is nothing new, it is precisely what occurs to the causal horizon in general relativity. New is however the fact that, here, the causal horizon is also the boundary of the classical universe, and its extension sets the size of masses and couplings. Therefore, not only time and length measurements undergo a relativity principle, but also those of masses and couplings. Besides the usual space-time relativity, one must consider also a "cosmological" relativity of mass/energy measurements. A consequence is the apparent acceleration of the rate of expansion of the universe, which in this scenario is not accelerated, although it appears to be, as a consequence of age/distance-related shifts in the spectrum of emitted light. In this way one can explain, and account, for a red shift typical of a matter dominated universe, and other deviations in the observed spectra.

Constraints usually considered "model killers" such as the nucleosynthesis or the Oklo bound are here completely reconsidered. In this scenario, they become harmless. The reason is that, since all mass scales and couplings run approximately as different powers of the age of the universe, constraints given by products and ratios of these quantities close to one almost do not scale with time. If a constraint is satisfied at a certain age, it is easily satisfied at any age.

Natural evolution and mutagenesis steps

Encouraged by the good agreement with the experimental results, we may take seriously this approach and investigate in this light also other domain of natural science. The fact that masses and couplings (and therefore any energy scale) depend on the age of the universe, irrelevant for instantaneous phenomena like particle scatterings, becomes relevant in all the natural phenomena which involve a long time span. It has therefore potential consequences also in the natural evolution of species. The Darwin's theory of evolution explains how, by natural selection, mutated species overwhelm and prevail over other ones, but says nothing about the biophysical mechanism that gives origin to mutations. Under the hypothesis that mutagenesis is triggered by some change in the molecular structure, one may argue that such a change is induced by the absorption of energy from natural radiation, in particular from the ultraviolet part of the spectrum of the light arriving to the earth from all over the surrounding universe. Energy absorption is then only possible if the energy corresponds to an absorption band of both the spectrum of the target, and the emission spectrum of a source. In absorption and emission spectra, frequencies are built in series over certain fundamental, time-dependent values. During the history of the universe, emission and absorption ground frequencies may vary in asynchronous way, implying absorption resonances at any time the emitted radiation hits an absorption frequency. The consequence is a "step-wise" evolution, in which there are phases of increased mutagenetic activity. Palaeontological observations seem to confirm this type of behaviour.

Quantum geometry and HTS

This theoretical framework is not just a quantum gravity scenario, in the sense of providing a theory of quantized gravitational interactions, something that extends the theory of elementary particles and interactions to include gravity, with consequences for high energy theory and cosmology. As it is clear from 1.0.1, we are here in the presence of a quantum theory of space geometries. This is much more: instead of seeking for a quantization of traditional degrees of freedom, here we proceeded in the opposite way: we first built a generalized theory of energy and space, and then, through the interpretation in terms of geometries, we obtained a physical theory potentially able to include any type of phenomenon. Geometries are energy clusters in the universe, energy packets corresponding to elementary particles and their interactions, organic molecules, and even any kind of materials, together with their physical properties. This scenario has therefore implications in many more domains than just elementary particle physics and cosmology. We have just seen how this includes mutagenesis. A further system, usually not associated with quantum gravity, are superconductors. The dependence of the critical temperature of high-temperature superconductors on the lattice structure of the material is a typical case in which the quantum properties of the geometry can not be neglected. Roughly speaking, quantum geometry results in producing a geometry-dependent Planck constant: $h \to h(g_{\mu\nu})$, or if one prefers $h(R)$, where R is the curvature. On the other hand, superconductivity is a matter of quantum delocalization of wavefunctions. It is therefore reasonable to expect a dependence of the critical temperature on the geometry of the crystal, such that the more the lattice structure is complex, i.e. less smooth, the higher is the quantum delocalization, and therefore also the critical superconduction temperature. This relation can be successfully investigated. What we obtain is a remarkable agreement between our prediction of the critical superconduction temperature, as derived from the analysis of the geometry of the crystal lattice of a superconductor, and the experimentally observed one. According to our analysis, high-temperature superconductivity turns out therefore to rely on the very same BCS mechanism as the low temperature one, the difference being only given by a material-related different degree of quantum delocalization of wavefunctions, a pure quantum geometry effect.

The universe of codes

In this theoretical set up, we build a quantum-relativistic world starting from simple rules of counting and classifying logical sequences. The assignments of geometries can be viewed as assignments of units (that we interpret as energy units) to units that we interpret as units of space. Altogether, these maps from a one-dimensional discrete vector space to a multi-dimensional discrete vector space represent the set of all the binary assignments of the type 0, 1 of a certain amount of 1's to a certain string of 0's. They represent therefore the set of all the binary codes. We may call them codes of information. Viewed in this way, the universe is then the whole of information. In practice, what we do is to propose an *interpretation* of logical structures in terms of physics. What we see, and live in, and are part of, is a path (that we call history) through the whole of information. The choice of discrete quantities as the fundamental objects of this construction is based on the consideration that numerable is more fundamental than non-numerable: any other type of numbers is in fact derived and can be logically expressed in terms of integer numbers. The scenario we obtain is therefore by no means more restrictive than a construction on the continuum: it is indeed the most comprehensive and general one can think about.

In this context, it is natural to expect that number theory plays an important role. In particular, it is intriguing to investigate the role of prime numbers. The phase space of any process occurring in the universe has a multiplicative structure. To any geometry, or subset of geometry, is associated a weight given by an integer number. A question is whether it is possible to associate a weight to any integer number, namely if any integer can be viewed as the weight of a certain geometry. Since we are talking of volumes of finite symmetry groups, one is tempted to give a positive answer to this question: a number simply *defines* a group, i.e. a symmetry associated to a finite number of elements. A related question is then about the role of non-factorizable weights, corresponding to prime numbers. From the properties of the distribution of primes within the integers we know that their "density" within the factorizable groups scales logarithmically: $\rho \sim \pi(n)/n \sim \ln n$. The weakly coupled sectors of the physical theory, like the sector associated to the electromagnetic interaction, and the photon that parametrizes the expansion of the horizon in the direction orthogonal to the tangent of space, are characterized by a logarithmic weight, reflecting in the logarithmic expression of their coupling strength as compared to the scaling of the space coordinates, and the size of the universe. Since these sectors concern long range interactions, it is tempting to associate them to symmetries that cannot be split into products of local terms, but involve the whole space, connecting at once any region of space. Their weight would seem to correspond to weights associated to prime numbers. Is there indeed a correspondence of the two, namely, between long range interactions and prime numbers, and between short range / strong coupling and non-prime integers? The logarithmic behaviour of the weight of long range interactions on one side, and the parallel logarithmic density of primes within integers seem to suggest the existence of a deep relation. Is then the duality of behaviours "logarithmic" versus "power-like" also the relation between perturbative and non-perturbative regime, weak versus strong coupling behaviour? This would open up a new perspective to the investigation of physical phenomena, and to the identification of elementary structures. Perhaps prime numbers play a fundamental role in a non-perturbative approach to interacting theory, as indicates also the short analysis on the recursiveness of certain structures at different energy scales that we present at the end of this work.

The chapters of this work

In **Chapter 2:** ***"A physical universe from the universe of codes"*** we start by setting up our framework, without assuming quantum mechanics or relativity as fundamental requirements of a basic physical construction. We introduce assignments of units to a discrete vector space. We discuss symmetries and compute the entropy for spheres of any dimension, to obtain that the favoured space is a 3-sphere. We introduce the time ordering as an ordering through increasing total energy, and the partition function of the universe of discrete geometries. Subsequently, we derive the Heisenberg's uncertainty, and discuss the type of dynamics this implies. We discuss then the speed of expansion of the universe, deriving the bound on the speed of propagation of coherent information, and the Lorentz boost in terms of transformation of entropies. The latter is the natural representation of coordinate transformations in a scenario in which also space is quantized. Once recognized that we are in the presence of a generalized quantum gravity theoretical framework, we discuss black holes, to conclude that these objects do not exist as we imagine them according to their classical formulation: going closer and closer to the Schwarzschild radius the classical notion of space is completely lost. The only such an object in the universe is the universe itself.

In **Chapter 3:** ***"The superstring representation of the universe of codes"*** we consider the large-energy limit, in which the discrete universe can be approximated by a continuous description of space and its geometry. In this limit, the physical scenario encoded in 1.0.1 naturally leads to a parametrization in terms of quantum superstrings. We introduce the concept of entropy, and entropy-weighted sum, in the phase space of the string constructions. We discuss the relation between the non-perturbative formulation and the representation of space in the perturbative constructions of the string. In particular, the perturbative limit is important because it allows to identify the spectrum of free particles. We introduce the concept of mass of elementary particles and fields, which in this scenario is related to that of coupling. We discuss how these quantities are related to volumes in the phase space, and how they are computed as functions of the age of the universe. Particular attention is devoted here to the strong interaction in general terms, to the reason and meaning of the existence of a strong force, besides an (electro-)weak one, discussing how its existence is necessarily required and implied by the coupling with gravity. We also compute the only mass eigenvalue of the strongly coupled universe. This will be associated to a state neutral (i.e. a bound state, a singlet) for all the interactions, apart from the gravitational one (in Chapter 4 this will turn out useful in order to derive the neutron mass). We discuss then how in the field theory limit the entropy-weighted sum 1.0.1 reduces to the Feynman path integral. The last part of the chapter contains a general discussion of the phenomenon of resonance in its various aspects, and how it arises as another consequence of the only rationale ruling the universe in this theoretical framework, namely the entropy in the phase space of all the configurations. In particular, we consider the physics of particle colliders, and the resonance peaks in the proton-antiproton high energy collisions.

Chapter 4: ***"The spectrum of the universe of codes"*** is the most technical and extended chapter of this work. It contains the investigation of the properties of the elementary particles and fields as derived from the spectrum of the string representation of this scenario. We start by investigating the string space through orbifold constructions. Through this, we obtain information about the pattern of progressive symmetry reduction, from which not only one obtains the spectrum of the propagating fields and particles, but also the ratios of their weights in the phase space, out of which all the mass ratios are computed. We derive the number of coordinates which remain untwisted and therefore free to expand, three, matching with the 3-sphere geometry of the geometric scenario. We pass then to a detailed computation of the mass of all the elementary particles. The latter turn out to correspond to the matter degrees of freedom of the Standard Model, and the gauge bosons of the weak interactions. We derive then the electromagnetic and weak coupling, and the coupling of the colour force. We compute also the mass of proton and neutron, related to the mass eingenvalue of the strongly coupled universe discussed in chapter 3. For all the couplings and masses we derive the dependence on the age of the universe, and the value at present time. A section is devoted to the discussion of the CKM matrix; in particular, we discuss how the experimental results on the flavour mixing and CP violation fit in this theoretical framework. We already said that in this scenario there is no Higgs field; the last section is devoted to a discussion of the S-dual, strong coupling phase of the electromagnetic interactions, which precisely predicts the occurrence of a series of resonances around 125 GeV.

In **Chapter 5:** ***"Cosmology"*** we discuss the cosmological evolution of the geometry as it results from the string computation. If on one hand this confirms the non-accelerated expansion of a 3-sphere, the analysis through the string representation, and therefore the results acquired from the investigation of the time scaling of masses and couplings of the elementary particles, allow to understand why such an expansion appears to be accelerated, in the form typical of a matter dominated universe. We discuss then the cosmic background radiation, and the issues of dark matter and cosmological constraints like the nucleosynthesis and the Oklo bound.

In **Chapter 6:** ***"The phases of the natural evolution"*** are discussed the implications of time-evolving energy scales on the steps of natural evolution. In particular, the phases of the evolution of primates and the Paleozoic-Mesozoic-Cenozoic sequence are plotted against the series of neighbouring resonances obtained under the hypothesis that mutagenesis is induced by the absorption of radiation from some natural source.

In **Chapter 7:** ***"High-temperature superconductivity"*** we address the problem of high-temperature superconductivity. We discuss how in our quantum geometry scenario an effective Planck constant is produced, that depends on the geometric properties of the material. As a consequence, also the threshold of the transition between classic and quantum regime depends on the geometry. Once established the relation between geometry and degree of quantum delocalization of wavefunctions, we obtain a recipe to predict the ratio of critical temperatures for a wide class of superconducting materials of which we know the crystal structure, finding a remarkable agreement with the temperatures measured experimentally.

In **Chapter 8:** ***"Prime numbers and the structures of the universe"*** we discuss how prime numbers can be a base of elementary blocks that in certain cases, and for a certain type of physical problems, could substitute the concept of asymptotically free state in a non-perturbatively interacting theory. We investigate then the relation between the scaling of primes within the natural numbers, and the analogous scaling of the couplings of long range forces, as compared to those of short range.

2 A physical universe from the universe of codes

2.1 The set-up

2.1.1 Distributing binary information

Consider a generic vector space, consisting of the Cartesian product of $M_1^{p_1} \times \ldots \times M_i^{p_i} \ldots \times M_n^{p_n}$ "elementary cells". Since an elementary, "unit" cell is basically adimensional, it makes sense to measure the volume of this p-dimensional space, $p = \sum_i^n p_i$, in terms of unit cells: $V = M_1^{p_i} \times \ldots \times M_n^{p_n}$. Although with the same volume, from the point of view of the combination of cells and attributes this space is deeply different from a one-dimensional space with V cells. To such a space we can assign, in the sense of "distribute", N "elementary" attributes, $N \leq V$. For the time being, we consider all M_i finite, so that the volume V is finite. This will turn out to be a regularization: at the end of the game this condition will be relaxed by taking the limit $M_i \to \infty$ for every i. On the other hand, N has always to be considered finite. In view of these considerations, it is therefore possible to assume that $M_i \gg N$, $\forall i$. What are these attributes? Cells, simply cells: our space is simply a mathematical structure of cells, and cells that we attribute to cells in certain positions. In this way, we are constructing a discrete function, an "assignment" $x = f(y)$, where y runs in the "attributes" and x belongs to the p-dimensional space. We define the phase space $\{\Psi(N)\}$ as the space of the assignments, i.e. the "maps" Ψ:

$$\Psi : N \to \prod_i M_i^{p_i}, \qquad M_i \geq N. \tag{2.1.1}$$

For large M_i and N, we can approximate the discrete degrees of freedom with continuous coordinates: $M_i \to r_i$, $N \to r$, with $r \ll r_i$ $\forall i$. We have therefore a continuous map $\Psi : y \in \{R\} \to \vec{x} \in \{R^p\}$ from a one-dimensional space of volume r to a p-dimensional space of volume $\prod r_i^{p_i}$.

The assignments 2.1.1 are basically assignments of binary codes. However, if we call N "total energy", and the M "space coordinates", the $\Psi(N)$ become assignments of geometries, and $\{\Psi(N)\}$ is the phase space of all the possible geometries at energy N. To stay general, let us call them "configurations". In order to appropriately compare configurations through the corresponding geometries, we may think of fixing a highest dimensionality of space, say P, fix a volume V_P of this P-dimensional space [1], and work with the subclass of configurations that correspond to spaces of dimension $p \leq P$, and volume smaller than V_P. In this way, all the geometries can be thought of as being embedded in a common, higher space. We will eventually let P and V_P go to infinity. We want to investigate what is the entropy of a certain configuration in this phase space. An important observation is that *there do not exist two configurations with the same entropy*: if they have the same entropy, they are perceived as the same configuration. The reason is that we have a combinatorial problem, and, at fixed N, the volume of occupation in the phase space is related to the symmetry group of the configuration. In practice, we classify configurations through statistics of combinations: a configuration corresponds to a certain combinatorial group. Now, discrete groups with the same volume, i.e. the same number of elements, are homeomorphic. This means that they describe the same configuration. Configurations and entropies are therefore in bijection with discrete groups, and this removes the degeneracy. Different entropy = different occupation volume = different volume of the symmetry group; in practice this means that we have a different configuration.

[1] Indeed, $P \leq V$ because it does not make sense to speak of a space direction with no more than one space cell.

2.1.2 Flat and curved geometry

When we distribute occupation numbers along a discrete vector space, there is a priori no intrinsic geometry: it is just a matter of pure combinatorics. However, starting from two dimensions, and above, it is possible to consider curved space geometries. Trivially, in order for this to make sense, the space must contain at least one cell occupied, otherwise there is no way of distinguishing geometries in an empty space. In our set up, geometry is therefore not an intrinsic property of the target space, but it is related to the energy content. Indeed, the presence of occupied cells has deep implications on the occupation probability of the remaining cells, because the entropy of the configuration depends on its symmetries. It is therefore not irrelevant the relative position of occupied cells. The presence of energy and its distribution determines therefore the probability of finding energy somewhere else. As we will see, we will eventually interpret energy clusters as objects such as particles etc.... Trajectories in space will be ruled by an entropy law. The presence of energy determines therefore dynamics and trajectories: the entropy of an energy cluster in empty space is different from its entropy in presence of energy somewhere else. In practice, to speak according to concepts familiar from general relativity, the presence of energy curves the space. The weight of a configuration. i.e. a distribution of energy in space, is related to the symmetries of the energy-determined geometry. For space dimensions higher than one, the presence of energy tells us that we are always in presence of a curved space. In this set up, there is no such a thing like a flat space, outside of the simple, trivial case of $N = 0$ and/or $d = 1$. The most entropic configurations are the "maximally symmetric" ones, i.e. those that look like spheres. We will therefore first consider the contribution to 1.0.2 as due to the geometry of the sphere.

In the following, we are interested in the large M, large N behaviour. In our language we will switch therefore forth and back from the discrete to its approximation in the continuum. Moreover, since we are interested in the scaling properties, we will neglect precise numerical coefficients. The weight in the phase space will be given by the number of times a sphere can be formed by moving along the symmetries of its geometry, times the number of choices of the position of, say, its centre, in the whole space. Since we eventually are going to take the limit $V \to \infty$, we don't consider here this second contribution, which is going to produce an infinite factor, equal for each kind of geometry, for any finite amount of total energy N.

2.1.3 Entropy of spheres

Let us start by considering the entropy of a 3-sphere. Curving the space implies a non-trivial interpretation of the boundaries of the energy distribution, as seen from a higher-dimensional embedding space [2]. This in general enhances the amount of symmetry. In order to evaluate the weight, we first investigate what happens for small increments of N. This necessarily means that we work on the tangent space. Consider the "differential equation" (more properly, a finite difference equation) of the increase in the number of combinations when passing from m to $m+1$. Owing to the multiplicative structure of the phase space (composition of probabilities), expanding by one unit the radius, or equivalently the scale of all the coordinates, adds to the possibilities to form the configuration for any dimension of the sphere some more $\sim m+1$ (that we can also approximate with m, because we work at large m) times the probability of one cell times the weight of the configuration of the remaining m (respectively $m-1$) cells. Depending on the scale of energy as compared to the space scale (in familiar words, on the value of conversion units such as c and $\hbar$), in general the sphere will not be a portion of space fulfilled with energy, i.e. entirely consisting of cells each one occupied by a unit of energy (radius $m_{\text{fulfilled}} \sim N^{1/3}$, density $\sim N/m^3_{\text{fulfilled}} = N/N = 1$), but will be a "sparse" space of lower density. Along this space, moving by a step shorter than the distance between cells occupied by an energy

[2] Think for instance of the relation between any stereographic representation of a 2-sphere in the two-dimensional plane, as compared to its representation in three dimensions.

unit will not be a symmetry, because one moves to a "hole" of energy. It is not difficult to realize that the effective symmetry group will have a volume $\mathcal{V}$ that stays to the volume $\mathcal{V}_{\text{fulfilled}}$ of a fulfilled space in the same ratio as the respective energy densities, $\mathcal{V}/\mathcal{V}_{\text{fulfilled}} = \left(N/m^3\right)/1$. We must therefore normalize the computation of the scaling multiplying by the energy density while at the same time fixing the scale of energy units as compared to the space units, N/m, i.e. dividing by this last factor. Taking into account all these effects, we obtain the following scaling:

$$W(m+1)_3 \sim W(m)_3 \times (m+1)^3 \times \frac{N}{m^3} \times \frac{m}{N}\,. \tag{2.1.2}$$

The factor N/m^3 is the density of the 3-sphere, while m/N fixes the energy-to-space coordinate scale [3]. Expanding $W(m+1)$ on the left hand side of 2.1.2 as $W(m) + \Delta W(m)$, and neglecting on the r.h.s. corrections of order $1/m$, we can write it as:

$$\frac{\Delta W(m)_3}{W(m)_3} \simeq m\,. \tag{2.1.3}$$

Since we are interested in the behaviour at large m, we can approximate it with a continuous variable, $m \to x$, and approximate the finite difference equation with a differential one. Upon integration, we obtain:

$$S_3 \propto \ln W(m)_3 \sim \frac{1}{2} m^2\,. \tag{2.1.4}$$

Fixing the radius/energy scale to 1, i.e. setting $m = N$, implies that the energy density of the 3-sphere scales as $1/N^2$. We obtain in this way an equivalence between energy density and curvature R:

$$\rho_3(N) \sim \frac{1}{N^2} \cong \frac{1}{r^2} \sim R_{(3)}\,. \tag{2.1.5}$$

[3]Indeed, in 2.1.2 there should be one more factor: when we pass from radius m to $m+1$ while keeping N fixed, the configuration becomes less dense, and we loose a symmetry factor of the order of the ratio of the two densities: $[m/(m+1)]^3 \sim 1 + \mathcal{O}(1/m)$.

This is basically the Einstein's equation relating the curvature of space to the tensor expressing the energy density. Indeed, here this relation can be *assumed* to be the *physical description* of a sphere in three dimensions. In order to preserve good properties of reduction of spaces to subspaces ($S^p \to S^{p-1} \to \ldots \to S^2$), we must *impose* a generalization of the above relation as condition in a generic dimension $p \geq 2$ for having the geometry of a sphere [4]:

$$\rho_p(E) \sim \frac{N}{m^p_{(p)}} \cong \frac{1}{m^2_{(p)}} . \tag{2.1.6}$$

In two dimensions, 2.1.6 implies $N = 1$ (up to some numerical coefficient). This means that, although it is technically possible to distribute $N > 1$ energy units along a 2-sphere of radius $m > 1$, from a physical point of view these configurations do not describe a sphere. Indeed, if we think of embedding the 2-sphere in three flat dimensions, we can view the energy $E = N$ as the "gravitational charge" of a central force with Coulomb-like potential, the usual gravitational potential $V \sim M/R$, where R is the radius of a 2-sphere enclosing the region with mass M. According to the Gauss's theorem, the flux of the gravitational field through the 2-sphere is equal to the mass M, which in our case is the total energy. But the flux is $\Phi \sim M/R^2 \times R^2$, independent on the radius. The gravitational charge, i.e. the mass, or total energy, can therefore be thought of as being concentrated at the center of the sphere. In this discrete scenario at the center we have just one space cell, on which we can accommodate only one unit of energy. The only 2-sphere *in two dimensions* is therefore the one with total energy $N = 1$. More energy units produce other kinds of geometries. In dimension $p \geq 3$ equation 2.1.6 is solved by:

$$m_{(p)} \sim N^{\frac{1}{p-2}} \quad (< N \text{ for } p > 3) \,, \tag{2.1.7}$$

[4] We recall that we omit here p-dependent numerical coefficients which characterise the specific normalization of the curvature of a sphere in p dimensions, because we are interested in the scaling at generic N, and m, in particular in the scaling at large N.

and the equivalent of 2.1.2 reads:

$$W_p(m_{(p)}+1) \sim W_p(m_{(p)}) \times (m_{(p)}+1)^p \times \frac{N}{m^p_{(p)}}\,, \qquad (2.1.8)$$

where we omit the scale-fixing factor that was present in 2.1.2, because we want to refer all geometries to the units of the three-dimensional one. We are going to take this into account by inserting the condition for the p-sphere expressed in equation 2.1.6. We obtain:

$$W_p(m_{(p)}+1) \sim W_p(m_{(p)}) \times (m_{(p)}+1)^p \times \frac{1}{m^2_{(p)}}\,, \qquad (2.1.9)$$

which leads to the following finite difference equation:

$$\frac{\Delta W(m_{(p)})_p}{W(m_{(p)})_p} \approx m^{p-2}_{(p)}\,. \qquad (2.1.10)$$

This expression obviously reduces to 2.1.3 for $p = 3$. Proceeding as before, by transforming the finite difference equation into a differential one, and integrating, we obtain:

$$S_{(p\geq 2)} \propto \ln W(m_{(p)}) \sim \frac{1}{p-1}\, m^{p-1}_{(p)}\,, \qquad p \geq 3\,. \qquad (2.1.11)$$

This is the typical scaling law of the entropy of a p-dimensional black hole (see for instance [23]). From expression 2.1.11 and 2.1.7 we derive:

$$S_{(p\geq 3)}|_N \sim \frac{1}{p-1}\, m^{p-1}_{(p)} \sim \frac{1}{p-1}\, N^{\frac{p-1}{p-2}}\,. \qquad (2.1.12)$$

This is the part of entropy that is due to the intrinsic symmetry of the p-dimensional sphere. In order to compare them within the higher-dimensional embedding space, we must think of lower-dimensional spheres as embedded in subspaces of the higher dimensional space. At any time we increase by one unit the dimension of the embedding space, from p to $p+1$, to the intrinsic entropy we must add a term which accounts for the fact that the p-dimensional subspace can be embedded in the $p+1$ dimensional rotated by various possible angles. The possible rotations occur along the p axes of the embedded subspace. The weight of the p-dimensional sphere gains therefore a factor of order $\sim m_{(p)}^{p} \sim N^{\frac{p}{p-2}}$. The weights of the spheres stay therefore in ratios of order:

$$\frac{W_{(p)}(N)}{W_{(p+1)}(N)} \approx N^{\frac{p}{p-2}} \times \exp\left[\frac{1}{p-1} N^{\frac{p-1}{p-2}} - \frac{1}{p} N^{\frac{p}{p-1}}\right] . \qquad (2.1.13)$$

By increasing p, as a function of N they are therefore exponentially suppressed with respect to each other. In particular, the largest suppression factor occurs between the three-dimensional sphere and the higher ones. Below three dimensions we cannot speak of spheres out of a trivial, formal sense. In two dimensions expression 2.1.8 can still be integrated if we intend the configuration as a collection of N spheres of energy $= 1$. We obtain:

$$W_2(N) \approx \left[\mathrm{e}^1\right]^N = \mathrm{e}^N , \qquad (2.1.14)$$

which is exponentially suppressed as compared to the 3-sphere. For $p = 1$ we cannot have anymore a curved space. We expect therefore no exponential dependence of the weight on N, but rather a power-law relation. From a formal point of view, we can still plug in 2.1.8 a "curvature" $1/m^2$, and obtain $\Delta W_1 / W_1 \sim 1/N$, that integrates to $W_1 \sim N$. However, to be more precise, we must consider that we are no more working on a linearization, i.e. on the tangent space to a point of a curved space, and expression 2.1.8 can no more be approximated by a differential equation. There is therefore no exponentiation of the dependence on N, which remains of power-law type. In both the $p = 2$ and $p = 1$ cases, the weights are exponentially suppressed as compared to the three-dimensional sphere. All this allows us to conclude that:

At any energy N, the most entropic configuration is the one corresponding to the geometry of a 3-sphere. Since in any dimension the sphere is also the most entropic geometry, three dimensions are statistically "selected out" as the dominant space dimensionality.

2.1.4 The "time" ordering

A property of $\{\Psi(N)\}$ is that, if $N_1 < N_2$, $\forall\Psi(N_1) \in \{\Psi(N_1)\}$ $\exists\Psi'(N_2) \in \{\Psi(N_2)\}$ such that $\Psi'(N_2) \supsetneq \Psi(N_1)$, something that, with an abuse of language, we write as: $\{\Psi(N_2)\} \supset \{\Psi(N_1)\}$, $\forall\ N_1 < N_2$. It is therefore natural to introduce an ordering in the whole phase space, that we call a "time-ordering", through the identification of N with the time coordinate: $N \leftrightarrow t$. We call "history of the universe" the "path" $N \to \{\Psi(N)\}$. This ordering turns out to naturally correspond to our everyday concept of time-ordering. In our normal experience, the reason why we perceive a history basically consisting in a progress toward increasing time lies on the fact that higher times bear the "memory" of the past, lower times. The opposite is not true, because "future" configurations are not contained in those at lower, i.e. earlier, times. But in order to be able to say that an event B is the follow up of A, $A \neq B$ (time flow from $A \to B$), at the time we observe B we need to also know A. This precisely means $A \in \{\Psi(N_A)\}$ <u>and</u> $A \subset A' \in \{\Psi(N_B)\}$, which implies $\{\Psi(N_A)\} \subset \{\Psi(N_B)\}$ in the sense we specified above. Time reversal is not a symmetry of the system [5].

2.1.5 How does a shape of space arise

In this set-up configurations are basically identified by their symmetry group. Configurations that describe the same geometry, but are "rotated" with respect to each other as compared to an external reference frame, actually describe *the same* configuration. The reason is that there is no "external frame": reference points are defined through the intrinsic asymmetries of the configurations themselves. Reference points are introduced through asymmetries. Starting from the most entropic one, we progressively obtain all the less entropic configurations by "moving" away the more and more units of energy to form less and less symmetric configurations, also walking through different dimensions. In this way, one obtains a *tower of asymmetric configu-*

[5] Only by restricting to some subsets of physical phenomena one can approximate the description with a model symmetric under reversal of the time coordinate, at the price of neglecting what happens to the environment.

rations "stapled" on the point at which the first asymmetry has been introduced. This point is therefore a reference point, that we assume to be *the point of the observer.* The superposition of configurations does not produce therefore a uniform universe, but a kind of "spontaneous" breaking of any symmetry. From the property, stated on page 34, that at any time $\mathcal{T} \sim N$ there do not exist two inequivalent configurations with the same entropy, and from the fact that less entropic configurations possess a lower degree of symmetry, we obtain that:

• *At any time $\mathcal{T}$ the average appearance of the universe is that of a space in which* **all symmetries are broken.**

As there is no external frame, in this framework there is also no external observer: an observer is a "local inhomogeneity" of space, and necessarily belongs to the universe. The observer is only sensitive to its own configuration, in the sense that he "learns" about the full space only through the superposition of configurations he is made of, and their changes. For instance, he can perceive that the configurations of space of which he is built up change with time, and *interprets* these changes as due to the interaction with an environment.

2.1.6 Mean values and observables

At any time $\mathcal{T} \sim N$ in the "universe" given by $\{\Psi(N)\}$ the mean value of any observable quantity $\mathcal{O}$ is the sum of the contributions to $\mathcal{O}$ over all configurations Ψ, weighted according to their volume of occupation in the phase space:

$$< \mathcal{O} > \propto \sum_{\Psi(\mathcal{T})} W(\Psi)\, \mathcal{O}(\Psi)\,. \qquad (2.1.15)$$

We have written the symbol $\propto$ instead of $=$ because the sum on the r.h.s. is not normalized. The weights don't sum up to 1, and not even do they sum up to a finite number: in the infinite volume limit, they all diverge [6]. However, as we discussed in section 2.1.1, what matters

[6] As long as the volume, i.e. the total number of cells of the target space, for any dimension, is finite, there is only a finite number of ways one can distribute

is their relative ratio, which is finite because the infinite volume factor is factored out. In order to normalize mean values, we introduce a functional that works as "partition function", or "generating function" of the universe:

$$\mathcal{Z} \stackrel{\text{def}}{=} \sum_{\Psi(\mathcal{T})} W(\psi) \;=\; \sum_{\Psi(\mathcal{T})} \mathrm{e}^{S(\Psi)} \,. \tag{2.1.16}$$

The sum has to be intended as always performed at finite volume. In order to define mean values and observables, we must in fact always think in terms of finite space volume, a regularization condition to be eventually relaxed. The mean value of an observable can then be written as:

$$< \mathcal{O} > \stackrel{\text{def}}{\equiv} \lim_{V \to \infty} \frac{1}{\mathcal{Z}} \sum_{\Psi(\mathcal{T})} W(\Psi)\, \mathcal{O}(\Psi) \,. \tag{2.1.17}$$

Mean values therefore are defined through an averaging procedure in which the weight is normalized to the total weight of all the configurations, at any finite space volume V.

2.1.7 Summing up geometries

Owing to the exponential suppression of any weight of a non-three-dimensional geometry, the mean value of the energy density is basically

energy units. In the infinite volume limit, both the number of possibilities for the assignment of energy, and the number of possible dimensions, become infinite.

the one measured in three dimensions:

$$\begin{aligned}
\langle \rho(E) \rangle &= \frac{1}{\sum_{\Psi(N)} W(\Psi(N))|_{d=3} + \sum_{\Psi(N)} W(\Psi(N))|_{d\neq 3}} \\
&\quad \times \left(\sum_{\Psi(N)} W(\Psi(N))\rho(E)_{\Psi(N)}|_{d=3} \right. \\
&\qquad \left. + \sum_{\Psi(N)} W(\Psi(N))\rho(E)_{\Psi(N)}|_{d\neq 3} \right) \\
&= \frac{\sum_{\Psi(N)} W(\Psi(N))\rho(E)_{\Psi(N)}|_{d=3} + \mathcal{O}\left(\mathrm{e}^{-N}\right)}{\sum_{\Psi(N)} W(\Psi(N))|_{d=3} + \mathcal{O}\left(\mathrm{e}^{-N}\right)} \\
&\approx \frac{\sum_{\Psi(N)} W(\Psi(N))\rho(E)_{\Psi(N)}|_{d=3}}{\sum_{\Psi(N)} W(\Psi(N))|_{d=3}} \\
&\qquad + \mathcal{O}\left(\mathrm{e}^{-N}\right) \, . \qquad (2.1.18)
\end{aligned}$$

We can therefore concentrate our analysis on three dimensions. Since the larger contribution to the mean value of the energy density is provided by the 3-sphere, we write 2.1.18 as:

$$\begin{aligned}
\langle \rho(E)_N \rangle &= \langle \rho(E)_N \rangle|_{S^3} + [d=3 \text{ corrections}] + \mathcal{O}\left(\mathrm{e}^{-N}\right) \\
&\approx \frac{1}{N^2} + [d=3 \text{ corrections}] \, . \qquad (2.1.19)
\end{aligned}$$

Let us now consider the contribution of geometries less symmetric than the sphere. In order to see what is the order of reduction of weight produced by displacing one energy unit one step away from its position on the sphere, consider the following: moving one energy unit by one step, one creates a "hole" in the former position and a neighbouring peak of energy. This deformation breaks the full geometric symmetry of the sphere. We assume that, as long as we depart by just one step away from the sphere, it is a reasonable approximation to consider that this leads to a reduction by a factor $\sim N^3$, the volume of the sphere. Since we have in total N units of energy, we have N equivalent possibilities of realizing this deformation. The overall reduction factor is therefore $\sim N \times 1/N^3 = 1/N^2$. We can figure out what is happening if we represent the configuration with the unit of energy displaced from A to A', a unit of space aside, as the superposition of the sphere plus the configuration in which the energy unit is removed from A (which therefore subtracts a certain amount of weight), plus the configuration in which the unit of energy is added in A'. In order to estimate the weight of this latter, we consider taking away a pair of units, AB, and add then the pair $A'B$. Indeed, the choice of B is irrelevant, as it is easy to see that the difference in weight between $[-(AB)+(A')B]$ and $[-(AC)+(A')C]$ only depends on the distance (AA'). Therefore, one can think of averaging over all the possible sums $[-(AB)+(A')B]$:

$$W(A') \; = \; \frac{1}{N^3} \times \sum_{i=1}^{N^3} \left[-W(AB_i) \; + \; W(A'B_i) \right] . \tag{2.1.20}$$

The weight of these configurations is simply the weight of a pair of units. We obtain therefore:

$$W(A') \; \sim \; \mathcal{O}(1) \, . \tag{2.1.21}$$

Since we can play this game with all the N units of energy of the sphere, we finally obtain:

$$\begin{aligned} W'(N+1) \; &= \; N \times \big(W_{(3)}(N) \; \times \; W(A') \big) \\ &\approx \; N \times \left(\frac{1}{N^3} \, \mathrm{e}^{N^2} \; \times \; \mathcal{O}(1) \right) \\ &\approx \; \mathcal{O} \left(\frac{\mathrm{e}^{N^2}}{N^2} \right) \, , \end{aligned} \tag{2.1.22}$$

thereby recovering as a result the previously estimated suppression factor of order $1/N^2$.

When we displace a second energy unit from the sphere, $B \to B'$, the distance and position of B' relative to A', the previously displaced one, are no more irrelevant in determining the weight, because we start from a situation of already broken symmetry. Since in the phase space we have $\sim N^3$ (the volume of the sphere) positions in which to equivalently realize the configuration $(A'B')$, a normalization $1/(N^3)$ factor is needed, leading to a further $1/(N^3)$ suppression factor in front of W'. Analogously to the previous case, the weight of the subtracted and added configurations results easier to compute if we think of subtracting from the sphere, and then adding back, one more unit of energy C, and averaging over C:

$$\begin{aligned}
W''(N+1)) &= N \times \{W'(N) \times W(A'B')\} \\
&\approxeq N \times \left\{ \frac{1}{N^3} W'(N) \right. \\
&\qquad \left. \times \frac{1}{N^3} \sum_{i=1}^{N^3} \left[-W(A'B(C_i)) + W(A'B'(C_i)) \right] \right\} \\
&\approx N \times \frac{1}{N^3} \left\{ \frac{N}{N^3} W(N-1) \times \mathcal{O}(1) \right\} \times \mathcal{O}(1) \\
&\approx \mathcal{O}\left(\frac{N^2}{(N^3)^2} \times \mathrm{e}^{N^2} \right) ,
\end{aligned} \tag{2.1.23}$$

that is:

$$W'' \approx \mathcal{O}\left(\frac{1}{N^2} W' \right) \approx \mathcal{O}\left(\left(\frac{1}{N^2} \right)^2 W \right) . \tag{2.1.24}$$

In these expressions, we have identified in the exponentials the numbers $N+1$, N and $N-1$, because in our derivation we re-normalize at any step to keep constant the radius of the sphere, even when subtracted of a small (as compared to N) number of points. Therefore, there is no $\exp -2N$ suppression factor coming from the squares in the exponential.

Similar considerations can be applied also to the further steps of reduction of symmetry, that therefore lead to a series of weight suppressions of order $\sim 1/N^2$. This is approximately true at least for the first steps of reduction. Going on displacing cells, there can occur also a partial restoration of symmetry. However, even in the case of reconstructing some product of spheres of smaller radius, something that can only happen once all the energy unit points have been so much displaced from their initial position on the sphere and rearranged, that we can no more use our approximation of keeping as reference point the weight $\exp N^2$ as the starting point of differential, power-like suppressions, the weight of the configuration is highly suppressed. For instance, in the case of a product of spheres $\prod_i S_i$ of radii $R_i \sim n_i$, $\sum_i n_i = N$, since $\sum_i n_i^2 = N^2 - 2\sum n_i n_{j\neq i}$ we have that the weight $W = \prod_i W_i \sim \mathrm{e}^{\sum n_i^2}$ is exponentially suppressed as compared to the weight of the unbroken 3-sphere.

We can view the operation of reducing the symmetry by progressively displacing energy by unit steps as a process of "soft breaking" tuned by an order parameter, in which each step breaks a piece of symmetry, leading to a suppression of the weight by at least a factor of order $1/N^2$. Summing up all the contributions leads to a correction which is of the order of the sum of an (almost) geometric series of ratio $1/N^2$. Similar arguments can be applied to $D \neq 3$, to conclude that expression 2.1.19 receives all in all a correction of order $1/N^2$. This result is remarkable. As we will discuss, the main contribution to the geometry of the universe, the one given by the most entropic configuration, can be viewed as the classical, purely geometrical contribution, whereas those given by the other, less entropic geometries, can be considered contributions to the quantum geometry of the universe. From 2.1.19 we see that not only the three-dimensional term dominates over all other ones, but that it is reasonable to assume that *the universe looks mostly like three-dimensional*, indeed mostly like a 3-sphere. This property becomes stronger and stronger as time goes by (increasing N). The d=3 corrections of expression 2.1.19 are roughly of order $1/N^2$ as compared to the first term:

$$\langle \rho(E)_N \rangle \approx \langle \rho(E)_N \rangle_{S^3} \left[1 + \mathcal{O}\left(\frac{1}{N^2} \right) \right] . \tag{2.1.25}$$

In general,

$$\sum_{\psi} W_{\psi}(N) = W(N)_{S^3} \left[1 + \mathcal{O}\left(\frac{1}{N^2} \right) + \mathcal{O}\left(\frac{1}{N^4} \right) + \ldots \right] . \tag{2.1.26}$$

From the fact that the maximal entropy is the one of a 3-sphere, and scales as $S_{(3)} \sim N^2$, we derive also that the ratio of the overall weight of the configurations at time $N-1$, normalized to the weight at time N, is of order:

$$W(N-1) \approx W(N)\, \mathrm{e}^{-2N} . \tag{2.1.27}$$

At any time, the contribution of past times is therefore negligible as compared to the one of the configurations at the actual time. This tells us that instead of 2.1.16 we could as well define the partition function of the universe at "time" $\mathcal{E}$ as the sum over all the configurations at past time/energy E up to $\mathcal{E}$:

$$\mathcal{Z}_{\mathcal{E}} = \sum_{\psi(E \leq \mathcal{E})} \mathrm{e}^{S(\psi)} . \tag{2.1.28}$$

2.2 The uncertainty principle

According to 2.1.17, quantities which are measurable by an observer living in three dimensions do not receive contribution only from the configurations of extremal or near to extremal entropy: all the possible configurations at a certain time contribute. Let us consider what does in practice means measuring the energy involved in a certain experiment. A measurement is the detection of the changes occurring in the shape, or geometry, of a certain subregion of the universe. Since the only thing one can do is detecting changes, it therefore necessarily implies a certain duration in time. The first thing one can think of measuring is the energy of the universe itself (for instance by measuring its curvature). For this, one needs a time long as much as the age of the universe itself: $\Delta t = N$. From expressions 2.1.25, 2.1.26 one can see that the corrections to the ground value $\langle E \rangle_0 = N$ are of order $1/N^2$, giving:

$$\begin{aligned} \langle E \rangle &= N + N \times \left[\mathcal{O}\left(\frac{1}{N^2} \right) + \text{higher orders} \right] \\ &\geq N + \frac{N}{N^2} \,. \end{aligned} \tag{2.2.1}$$

Considering that not only $\langle E \rangle_0 = N$ but also $\Delta t = N$, this expression can be written as:

$$\Delta \langle E \rangle \geq \frac{1}{\Delta t} \,. \tag{2.2.2}$$

Let us now consider a local experiment. This involves just a subregion of the whole universe. In general, the geometry is produced by a staple of configurations locally very far from the simple, "empty" space characterising the ground geometry of universe, the 3-sphere. Accordingly, also its entropy is very suppressed as compared to the entropy of the sphere. On the other hand, when one measures a local experiment, the contribution of the rest of the universe, and the ground contributions to the geometry, are implicitly subtracted from the description and the measurements. This operation is made possible by the properties of factorization of the phase space. Owing to its multiplicative structure, we can think of the higher order correction to the ground shape of the experiment as being produced by local subsets of the whole geometries of the universe, for which we can apply the result 2.1.26, this time restricted to a sub-factor of the weight, corresponding to the local region of space of radius N' ($N' < N$) large as much as the duration of the measurement: $N' = \Delta t'$. Reasoning as before, we can conclude once again that, during the time interval $\Delta t'$, the corrections to the energy of the experiment (i.e. of its geometry) are *at least* of order of the corrections to the energy of a small universe of radius $\Delta t'$:

$$\Delta \langle E' \rangle \geq \frac{1}{\Delta t'} \,. \tag{2.2.3}$$

Notice that, in this case, E' does not need to be itself large as much as the inverse of the time interval $\Delta t'$. What is large as much as the inverse of the elapsed time is the minimal correction to the energy of the small region, which, once subtracted of the geometry of the experiment, can be compared to a small, “empty” universe. This relation can be written as:

$$\Delta E \, \Delta t \gtrsim 1 \,. \tag{2.2.4}$$

This expression must be compared with the time-energy Heisenberg's uncertainty relation (introducing the Planck constant is here just a matter of introducing units enabling to measure energies in terms of time). In our case, it directly proceeds from the very definition of the physical set up, i.e. from the fact that the evolution of the universe is a history through superpositions of an infinite number of geometries. The bound to an experimental access to the universe corresponds to the limit within which such a universe is in itself defined. In this set up,

• *it is not possible to go beyond the uncertainty principle's bound with the precision in the measurements, because this bound corresponds to the precision with which the quantities to be measured themselves are defined.*

2.3 Deterministic or probabilistic physics?

The scenario implied by the sum 2.1.16 is neither probabilistic in the usual sense of quantum mechanics, nor deterministic according to the usual meaning of causality. Rather, it is ”determined” at any time by the partition function. The universe at time $N' \sim \mathcal{T}' = \mathcal{T} + \delta\mathcal{T} \sim N + 1$ is not obtained by running forward, possibly through equations of motion, the configurations at time $N \sim \mathcal{T}$, it is not their “continuation”: it is given by the weighted sum of all the configurations at time $\mathcal{T} + \delta\mathcal{T}$, as the universe at time $\mathcal{T}$ was given by the weighted sum of all the configurations at time $\mathcal{T}$. In the large N limit, we can speak of “continuous time evolution” only in the sense that for a small change of time, the dominant configurations correspond to geometries that don't differ that much from those at previous time. With a certain approximation we can therefore speak of evolution in the ordinary sense of (differential, or difference) time equations. Owing to the fact that at any time the appearance of the universe is mostly determined by the most entropic configurations, in the average

- *the dynamics of the evolution of the system is of entropic type.*

On the other hand, a full knowledge of the infinite terms of 2.1.16 is impossible, and, owing to the fact that configurations in any dimensions are taken into account, also ill-defined. From this point of view, the probabilistic interpretation of the Heisenberg's uncertainty given in quantum mechanics seems a viable way of parametrizing the unknown, reintroducing thereby a certain degree of predictability and calculability. This is also the case of systems in which the asymmetries are "hidden" below the threshold of the uncertainty 2.2.4, and produce therefore the impression of equal probability of equivalent situations, like the two possible paths of an electron in the double slit experiment: being able to predict the details of an event, such as for instance the precise position each electron will hit on the plate, and in which sequence, requires knowing the function "entropy" for an infinite number of configurations, corresponding to any space dimensionality at fixed $\mathcal{T} \approx N$, for any time $\mathcal{T}$ the experiment runs on. Clearly, no computer or human being can do that. If on the other hand we content ourselves with an approximate predictive power, we can roughly reduce physical situations to certain ideal schemes, such as for instance "the symmetric double slit" problem. Of course, from a theoretical point of view we lose the possibility of predicting the position the first electron will hit the target (something anyway practically impossible to do), but we gain, at the price of introducing symmetries and therefore also concepts like "probability amplitudes", the capability of predicting with a good degree of precision the shape an entire beam of electrons will draw on the plate. We give up with the "shortest scale", and we concern ourselves only with an "intermediate scale", larger than the point-like one, shorter than the full history of the universe itself. The interference pattern arises as the dominant mean configuration, as seen through the rough lens of this "intermediate" scale. In this scenario, quantum de-coherence is "built-in" in 2.1.16.

2.4 Relativity

As we discussed in sections 2.1.3–2.1.7, although the volume of the target spaces of the maps $\Psi(N)$ is eventually to be considered infinite, $V \to \infty$, at any finite time the dominant configuration of the universe corresponds to a 3-sphere of radius $N \sim \mathcal{T}$. Next to this, there is a staple of many "almost spherical", three-dimensional configurations that, in the superposition, give rise to a space with energy clusters. In the sum 2.1.16 there are also configurations which correspond to a geometry not bounded within a region of radius $N \sim \mathcal{T}$, nor three-dimensional. Indeed, for any V, there are configurations which "fulfill" the volume. They contribute in the form of quantum perturbations, all of them falling under the "cover" of the uncertainty principle, and being therefore related to what we interpret as the quantum nature of physical phenomena. All this can be interpreted in the following way: at any finite time $\mathcal{T}$ we have a universe which is infinitely extended, but that can be organized by separating it into a "classical part", with a geometry looking like the interior of a black hole, with a horizon placed at distance $\propto \mathcal{T}$, and a quantum part, which accounts for the contribution of any other kind of configurations. Only the classical part can be reduced to the ordinary geometric interpretation of space extended only up to a distance $\propto \mathcal{T}$. In this perspective,

- *the space "outside" the horizon is infinitely extended, but it contributes to the perception of a classical observer and to the values of the observables defined in the three-dimensional classical space only through the uncertainty of mean values, accounted for by the Heisenberg's uncertainty.*

In the following we want to see how in this universe Einstein's special (and general) relativity are implied as a particular limit, in which one considers just the classical part of space.

2.4.1 From the speed of expansion of the universe to a maximal speed for the propagation of information

The classical space at time N corresponds to a universe of radius $\sim N$, with total energy N. It expands at speed 1. Indeed, we can introduce a factor of conversion from time to space, c, and say that, by choice of units, we set the speed of expansion to be $c = 1$ (in an obvious way, also the conversion between units of space, and time, on one side, and energy on the other side, is here "by default" set to one, but it can be called h). We want to see how this is also the maximal speed for the propagation of information within the classical space. It is important to stress that all this refers only to the classical space as we have defined it, because only in this sense we can say that the universe is three dimensional: the sum 2.1.16 contains in fact also configurations that, through the time flow, can be interpreted as "tachyonic", along with configurations in which it is not even clear what is the meaning of speed of propagating information in itself, as there is no recognizable information at all, at least in the sense we usually intend it. Indeed, when we say we get information about, say, the motion of a particle, or a photon, we intend to speak of a non-dispersive wave packet, so that we can say we observe a particle, or photon, that remains particle, or photon, along its motion [7] (the existence, in the scenario implied by 2.1.16, of structures of this kind, namely of wave packets that behave like massive particles, or massless photons etc., will be investigated in the next chapters). Let's consider the simplified case of a universe at time N containing only one such a wave packet [8], as illustrated in figure 2.1, where it is represented by the shadowed cells, and the space is reduced to two dimensions.

[7] Like a particle, also a physical photon, or any other field, is not a pure plane wave but something localized, therefore a superposition of waves, a wave packet.

[8] We may think to consider only a portion of the universe, where only such a wave packet is present.

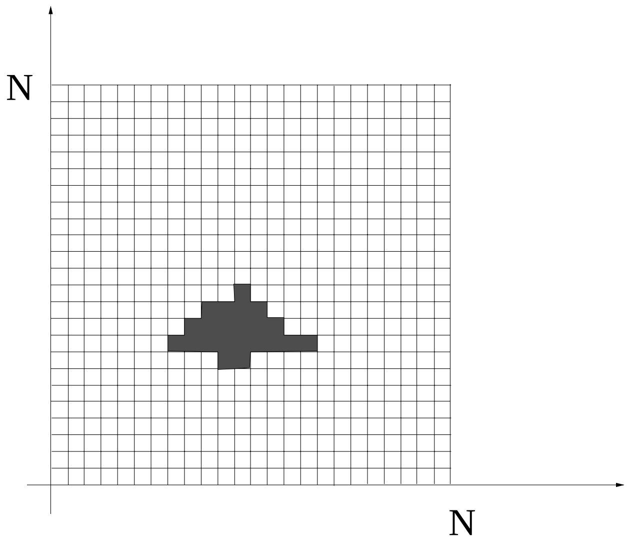


Figure 2.1: The initial position of an energy packet at time N.

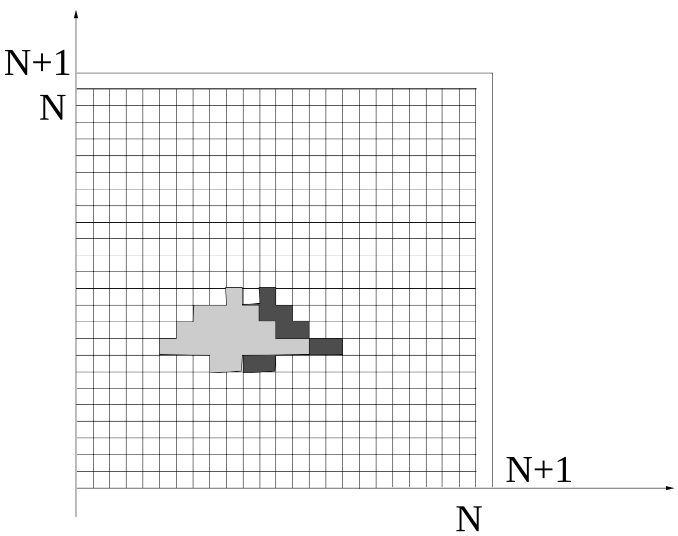


Figure 2.2: The energy packet at time N+1, displaced by two cells.

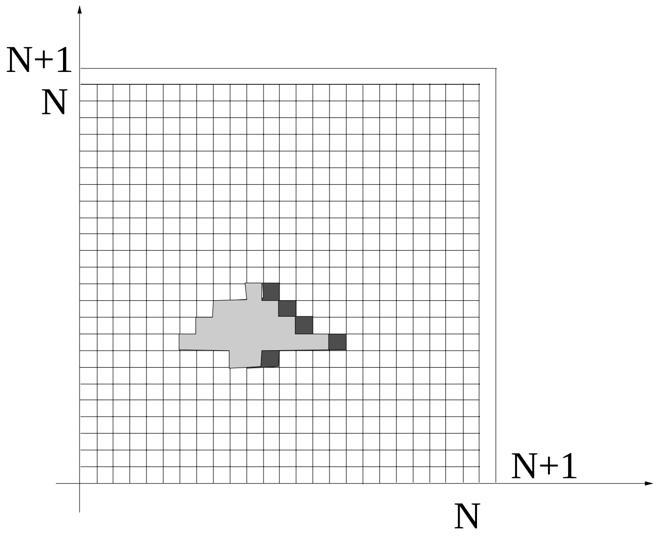


Figure 2.3: The same energy packet at time N+1, displaced by just one cell.

Consider now the evolution at the subsequent instant of time, namely after having progressed by a unit of time. Adding one point, $N \to N + 1$, produces an average geometry of a three sphere of radius $N + 1$ instead of N. In the average, it is therefore like having added $4\pi N^2$ "points", or unit cells. Remember that we work always with an infinite number of cells in an unspecified number of dimensions; when we talk of universe in three dimensions within a region of a certain radius, we just talk of the dominant geometry. Let's suppose the position of the wave packet jumps, back or forth, by two-cells steps, as illustrated in figure 2.2. Namely, as the time, and consequently also the radius of the universe, progresses by one unit, the packet moves at higher speed, jumping by two units. Compare this case with the case in which the packet jumps by just one unit, as in figure 2.3. The entropy of this latter configuration, intermediate between the first and second one, cannot be very different from the one of the second configuration, figure 2.2, in which the packet jumps by two steps, because that was supposed to be the dominant configuration at time $N+1$, and therefore the one of maximal entropy. Indeed, by "continuity" it must interpolate between step 2 and the configuration at time N, that was also supposed to be a configuration of maximal entropy. Therefore, the actual appearance of the universe at time $N+1$ must be somehow a superposition of the configurations 2 and 3, thereby contradicting our hypothesis that the wave packet is non-dispersive [9]. Therefore, the wave packet cannot jump by two steps, and we conclude that the maximal speed allowed is that of expansion of the radius of the universe itself, namely, c.

It is too early here to discuss the actual existence in this scenario of degrees of freedom that can be interpreted as photons. In order to do this we must pass to a representation on the continuum, where, as we will discuss in the next chapter, it corresponds to a string scenario. Here we just anticipate that, according to this theoretical framework, the reason why we have a universal bound on the speed of light is that light carries what we call classical information. Information about whatever kind of event tells about a change of average entropy of the observed system, of the observer, and also what surrounds and connects them. The rate of transfer/propagation of information is therefore strictly related to the rate of variation of entropy. Variation of entropy is what gives the measure of time progress in the universe. Any carrier of information that "jumps" steps of the evolution of the universe, going faster than its rate of entropy variation, becomes therefore dispersive, loosing information during its propagation. Light must therefore propagate at most at the rate of expansion of space-time (i.e. of the universe itself), what is usually called the speed of light in vacuum.

2.4.2 The Lorentz boost

Let's now consider physical systems that can be identified as massive particles, i.e. let us assume that there are local superpositions of configurations which are interpreted as travelling at speeds always lower than c. Since the phase space has a multiplicative structure, and

[9] If it was dispersive, it would be something like a particle that, during its motion, "dissolves", and therefore we cannot anymore trace as a particle. It would be just a "vacuum fluctuation" without true motion, something that does not carry any information in the classical sense.

entropy is the logarithm of the volume of occupation in this space, for each such a system it is possible to separate the entropy into the sum of an internal, "rest" entropy, and an external, "kinetic" entropy. The first one refers to the structure of the system in itself, that can be a particle or an entire laboratory (a point-like particle is an extended object of which we neglect the geometric structure). The second one refers to the relation/interaction of this system with the environment, the external world: its motion, the accelerations and external forces it experiences, etc.

Let us for a moment abstract from the fact that the actual configuration of the universe implied at any time by 2.1.16 describes a curved space. In other words, let's neglect the so called "cosmological term". This approximation can make sense at large N, as is the case of the present-day physics. This means, at large age of the universe [10]. Let us also assume we can just focus our attention on two observers sitting on two <u>inertial</u> frames, A and A', moving at relative speed v, neglecting everything else. For what said above, $v < 1$. An experiment is the measurement of some event; owing to the fact that happening of something means changing of entropy and therefore is equivalent to a time progress, it is perceived as having taken place during a certain interval of time. Let us consider an experiment, i.e. the detection of some event, taking place in the co-moving frame of A', as reported by both the observer at rest in A, and the one at rest in A' (from now on we will indicate with A, and A', indifferently the frame and the respective observer as well). Let's assume we can neglect the space distance separating the two observers, or suppose there is no distance between them [11]. For what we said above, such a detection amounts in observing the increase of entropy corresponding to the occurring of the event, as seen from A, and from A' itself. Since we are talking

[10] To make contact with ordinary physics, consider that, once expressed in units in which the Planck constant and the speed of light are 1, the present age of the universe is estimated to be of order 10^{31}, and the cosmological constant of order $\Lambda \sim 10^{-61}$. It is precisely its smallness what historically allowed to introduce special relativity and Lorentz boosts before addressing the problem of the cosmological constant.

[11] In our scenario, huge (=cosmic) distances have effect on the measurement of masses and couplings.

of the same event, the *overall* change of entropy will be the same for both A and A'. One would think there is an "absolute" time interval, related to the evolution of the universe corresponding to the change of entropy due to the event under consideration. However, the story is rather different as soon as we consider *time measurements* of this event, as reported by the two observers, A and A'. The reason is that the two observers will in general attribute in a different way what amount of entropy change has to be considered a change of entropy of the "internal" system, and which amount refers to an "external" change. Proper time measurements have to do with the *internal* change of entropy. For instance, consider the entropy of all the configurations contributing to form, say, a clock. The part of phase space describing the uniform motion of this clock will not be taken into account by an observer moving together with the clock, as it will not even be measurable. This part will however be considered by the other observer. Therefore, when reporting measurements of time intervals made by two clocks, one co-moving with A, and one seen by A to be at rest in A', owing to a different way of attributing elements within the configurations building up the system, between "internal" and "external", we will have in general two different time measurements. Let us indicate with ΔS the change of entropy as is observed by A. We can write:

$$\begin{aligned} \Delta S \, (\equiv \Delta S(A)) \;&=\; \Delta S(\text{internal} \,=\, \text{at rest}) \\ &\qquad + \Delta S(\text{external}) \qquad (2.4.1) \\ &=\; \Delta S(A') \,+\, \Delta S_{\text{Kinetic}}(A)\,, \qquad (2.4.2) \end{aligned}$$

with the identifications $\Delta S(\text{internal} \,=\, \text{at rest}) \equiv \Delta S(A')$ and $\Delta S(\text{external}) \equiv \Delta S_{\text{Kinetic}}(A)$. In section 2.1.3 we discussed how the entropy of a 3-sphere is proportional to $N^2 = E^2$. This is therefore also the entropy of the average, classical universe, that in the continuum limit, via the identification of total energy with time, can be written as:

$$S \;\propto\; (c\mathcal{T})^2 \,, \qquad (2.4.3)$$

where $\mathcal{T}$ is the age of the universe. This relation matches with the Hawking's expression of the entropy of a black hole of radius $r = c\mathcal{T}$

[24, 25]. It is not necessary to write explicitly the proportionality constant in (2.4.3), because we are eventually interested only in ratios of entropies. During the time of an event, Δt, the age of the universe passes from $\mathcal{T}$ to $\mathcal{T} + \Delta t$, and the variation of entropy, $\Delta S = S(\mathcal{T} + \Delta t) - S(\mathcal{T})$, is:

$$\Delta S \ \propto \ (c\Delta t)^2 \ + \ c^2 \mathcal{T}^2 \left(\frac{2\Delta t}{\mathcal{T}} \right) . \tag{2.4.4}$$

The first term corresponds to the entropy of a "small universe", the universe which is "created", or "opens up" around an observer during the time of the experiment, and embraces within its horizon the entire causal region about the event. The second term is a "cosmological" term, that couples the local physics to the history of the universe. The influence of this part of the universe does not manifest itself through elementary, classical causality relations within the duration of the event, but indirectly, through a (slow) time variation of physical parameters such as masses and couplings, (the time dependence of masses and couplings will be discussed in chapters 3 and 4). In the approximation of our abstraction to the rather ideal case of two inertial frames, we must neglect this part, concentrating the discussion to the local physics. In this case, each experiment must be considered as a "universe" in itself. Let's indicate with Δt the time interval as reported by A, and with $\Delta t'$ the time interval reported by A'. In units for which $c = 1$, and omitting the normalization constant common to all the expressions like 2.4.3 , we can therefore write:

$$\Delta S(A) \to \langle \Delta S(A) \rangle \approx (\Delta t)^2 \, , \tag{2.4.5}$$

whereas

$$\Delta S(A') \to \langle \Delta S(A') \rangle \approx (\Delta t')^2 \, , \tag{2.4.6}$$

and

$$\Delta S_{\mathrm{Kinetic}}(A) \ = \ (v \, \Delta t)^2 \, . \tag{2.4.7}$$

These expressions have the following interpretation. As seen from A, the total increase of entropy corresponds to the black hole-like entropy of a sphere of radius equivalent to the time duration of the experiment.

Since $v = c = 1$ is the maximal "classical" speed of propagation of information, all the classical information about the system is contained within the horizon set by the radius $c\Delta t = \Delta t$. However, when A attempts to refer this time measurement to what A' could observe, it knows that A' perceives itself at rest, and therefore it cannot include in the computation of entropy also the change in configuration due to its own motion (here it is essential that we consider inertial systems, i.e. constant motions). "A" separates therefore its measurement into two parts, the "internal one", namely the one involving changes that occur in the configuration as seen at rest by A' (a typical example is for instance a muon decay at rest in A'), and a part accounting for the changes in the configuration due to the very being A' in motion at speed v. If we subtract the internal changes, namely we think of the system at rest in A' as at a point without meaningful physics apart from its motion in space [12], the entire information about the change of entropy is contained in the "universe" given by the sphere enclosing the region of its displacement, $v^2(\Delta t)^2 \; = \; \Delta S_{\mathrm{Kinetic}}(A)$. In other words, once subtracted the internal physics, the system behaves, from the point of view of A, as a universe which expands at speed v, because the only thing that happens is the displacement itself, of a point otherwise fixed in the local universe (see figure 2.4). Inserting expressions 2.4.5–2.4.7 in 2.4.2 we obtain:

$$(\Delta t)^2 \; = \; \frac{(\Delta t')^2}{1 - v^2} \, , \tag{2.4.8}$$

that is:

$$\Delta t \; = \; \frac{\Delta t'}{\sqrt{1 - v^2}} \, . \tag{2.4.9}$$

The time interval as measured by A results to be longer by a factor $(\sqrt{1 - v^2})^{-1}$ than as measured by A'. In this argument the bound on the speed of information, and therefore of light, enters when we write the variation of entropy of the "local universe" as $\Delta S = (c\Delta t)^2$. If $c \to \infty$, namely, if within a finite interval of time an infinitely extended

[12] No internal physics means that we also neglect the contribution to the energy, and entropy, due to the mass.

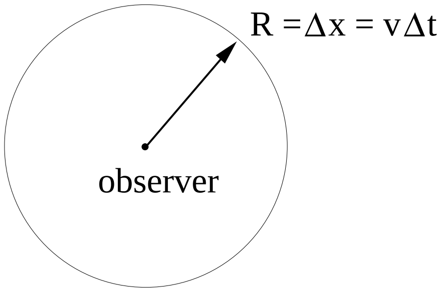


Figure 2.4: During a time Δt, the pure motion "creates" a universe with an horizon at distance $\Delta x = v\Delta t$ from the observer. As seen from the rest frame, this part of the physical system does not exist. The "classical" entropy of this region is given by the one of its dominant configuration, i.e. it corresponds to the entropy of a black hole of radius Δx.

causal region opens up around the experiment, both A and A' turn out to have access to the full information, and therefore $\Delta t = \Delta t'$. This means that they observe the same overall variation of entropy.

2.4.2.1 The space boost

In this framework we obtain in quite a natural way the Lorentz time boost. The reason is that, for us, the time evolution is directly related to the change of entropy, and we identify configurations (and geometries) through their entropy. The space length is somehow a derived quantity, and we expect also the space boost to be a secondary relation. Indeed, it can be easily derived from the time boost, once lengths and their measurements are properly defined. However, these quantities are less fundamental, because they are related to the classical concept of geometry. We could produce here an argument leading to the space boost. However, this would basically be a copy of the classical derivation within the framework of special relativity. The derivation of the time boost through entropy-based arguments opens instead new perspectives, allowing to better understand where relativity ends and quantum physics starts. Or, to better say, it provides us with an embedding of this problem into a scenario that contains both these aspects, relativity and quantization, as particular cases, to be dealt with as useful approximations.

2.4.3 General time coordinate transformation

Lorentz boosts are only a particular case of the general coordinate transformation, obtained within the context of general relativity; in that case the measure of time lengths is given by the time-time component of the metric tensor. In the absence of mixing with space boosts, i.e., with a diagonal metric, we have:

$$(ds)^2 \;=\; g_{00}(dt)^2 \,. \tag{2.4.10}$$

As the metric depends on the matter/energy content through the Einstein's equations:

$$\mathcal{R}_{\mu\nu} - \frac{1}{2} g_{\mu\nu}\mathcal{R} \;=\; 8\pi G_N T_{\mu\nu} \,, \tag{2.4.11}$$

g_{00} can be computed when we know the energy of the system. For instance, in the case of a particle of mass m moving at constant speed v (inertial motion), the energy, the "external" energy, is the kinetic energy $\frac{1}{2}mv^2$, and we recover the v^2-dependence of the Lorentz boost [13].

In the simple case of the previous section, we have considered the physical system of the wave packet as decomposed into a part experiencing an "internal" physics, and a part which corresponds to the point of view of the center of mass, that is, a part in which the complex internal physics is dealt with as a point-like particle. The Lorentz boost has been derived as the consequence of a transformation of entropies. Indeed, our coordinate transformation is based on the same physical grounds as the usual transformation of general relativity, based on a metric derived from the energy tensor. Let us consider the

[13] In the determination of the geometry, what matters here is not the full force experienced by the particle but the field in which the latter moves. The mass m therefore drops out from the expressions (see for instance [26]).

transformation from this point of view: although imprecise, the approach through the linear approximation helps to understand where things come from. In the linear approximation, where one keeps only the first two terms of the expansion of the square-root $\sqrt{1-v^2/c^2}$, the Lorentz boost can be obtained from an effective action in which in the Lagrangian appear the rest and the kinetic energy. These terms correspond to the two terms on the r.h.s. of equation 2.4.2. Entropy has in fact the dimension of an energy multiplied by a time [14]. Approximately, we can write:

$$\Delta S \;\simeq\; \Delta E \Delta t\,, \tag{2.4.12}$$

where ΔE is either the kinetic, or the rest energy. The linear version of the Lorentz boost is obtained by inserting in (2.4.12) the expressions $\Delta E_{rest} = m$ and $\Delta E_{kinetic} = \frac{1}{2}mv^2$. In this case, the linearization of entropies lies in the fact that we consider the mass a constant, instead of being the full energy of the "local universe" contained in a sphere of radius Δt, i.e. the energy (mass) of a black hole of radius Δt: $m = \Delta E = \Delta t/2$. In our theoretical framework, the general expression of the time coordinate transformation is:

$$(\Delta t')^2 \;=\; \langle \Delta S'(t) \rangle \;-\; \langle \Delta S'_{external}(t) \rangle\,. \tag{2.4.13}$$

Here $\Delta S'(t)$ is the total variation of entropy of the "primed" system as measured in the "unprimed" system of coordinates: $\langle \Delta S'(t) \rangle = (\Delta t)^2$. We can therefore write expression 2.4.13 as:

$$(\Delta t')^2 \;=\; [1 - \mathcal{G}(t)]\,(\Delta t)^2\,, \tag{2.4.14}$$

where:

$$\mathcal{G}(t) \;\stackrel{\text{def}}{=}\; \frac{\Delta S'_{external}(t)}{(\Delta t)^2}\,. \tag{2.4.15}$$

With reference to the ordinary metric tensor $g_{\mu\nu}$, we have:

$$\mathcal{G}(t) \;=\; 1 - g_{00}\,. \tag{2.4.16}$$

[14] By definition, $dS = dE/T$, where T is the temperature, and remember that in the conversion of thermodynamic formulas, the temperature is the inverse of time.

$\Delta S'_{external}(t)$ is the part of change of entropy of A' referred to by the observer A as something that does not belong to the rest frame of A'. It can be the non accelerated motion of A', as in the previous example, or more generally the presence of an external force that produces an acceleration. Notice that the coordinate transformation 2.4.14 starts with a constant term, 1: this corresponds to the rest entropy term expressed in the frame of the observer. For the observer, the new time metric is always expressed in terms of a deviation from the identity.

By construction, 2.4.15 is the ratio between the metric in the system which is observed and the metric in the system of the observer. From such a coordinate transformation we can pass to the metric of space-time itself, provided we consider the coordinate transformation between the metric g' of a point in space-time, and the metric of an observer which lies on a flat reference frame, whose metric is expressed in flat coordinates. We have then:

$$1 - \mathcal{G}(t) \; = \; \frac{g^{(\prime)}_{00}}{g^{(0)}_{00} \; = \eta_{00} = 1} \, . \tag{2.4.17}$$

As soon as this has been clarified, we can drop out the denominator and we rename the primed metric as the metric tout court.

2.4.4 General relativity

The set $\{ \Psi(\mathcal{T}) \}_{\mathcal{T} \geq 1}$ corresponds to the history of a universe consisting of evolving geometries in the most general sense. We have seen that this universe embeds the uncertainty principle and, in appropriate limits, special relativity. We may ask whether also general relativity is accounted for. We cannot expect 2.1.16 to exactly reproduce this theory, because it contains more. Indeed, we will see that, in order to investigate the spectrum of its physical content, the best translation in terms of local fields and interactions is provided by string theory, when appropriately embedded. Nevertheless, as we did for the ground principles of quantum mechanics, and the Lorentz boost, we want to discuss here in what terms general relativity is indeed contained in this scenario. By construction, it is the distribution of energy what determines the geometry. However, we cannot speak of "motion along geodesics", as we have rather here an evolution of geometries, ruled by an entropic principle: at any time the shape of the universe is dominated by the staple of its most entropic configurations. At large time and when restricted to the most entropic ones, the evolution can be *approximated* by a continuous motion through the geometry of an expanding universe. What substitutes the motion along geodesics is here a stepwise evolution according to the principle of maximising entropy at any step. This is the closest generalization of the motion along geodesics: non-minimal distance paths are unfavoured as compared to minimal distance ones, because they don't maximise entropy. Therefore, although present in the sum 2.1.16, they give a suppressed contribution as compared to the minimal distance paths. Let us consider an example that makes this concretely understandable. Although this set up at any time accounts for configurations of the entire universe, let us consider, with a certain amount of abstraction, the non-realistic case of a "universe" consisting of just two spheres with radius N_1 and N_2 respectively, placed in A and B, at a certain distance from each other, as in figure 2.5. The overall entropy is roughly given by the product of the entropies of the single spheres, times a factor depending on their relative distance: the more the two spheres are far apart from each other, the lower is the number of possibilities we have to place this configuration within a certain volume of space, and consequently the smaller is this factor. Therefore, the system will evolve toward a configuration in which the two spheres come closer to each other, to the point they will "fuse" to form a sphere of radius $N_{1+2} = N_1 + N_2$, with entropy $\sim \exp(N_1 + N_2)^2 > \exp N_1^2 \times \exp N_2^2$. This is a sketch of how gravitational attraction works. Of course, in the real scenario the geometry is far more complicated and, as we will see, there are details of the quantum part of the geometry that imply a description in terms of more degrees of freedom than just the mass (indeed, we will recover the spectrum of all the fundamental interactions, but, for the time being, let us neglect these aspects and just think of this simplified system). Let us now compare the two situations of figure 2.5, where the sphere 2, placed in B, moves toward sphere 1 placed in A,

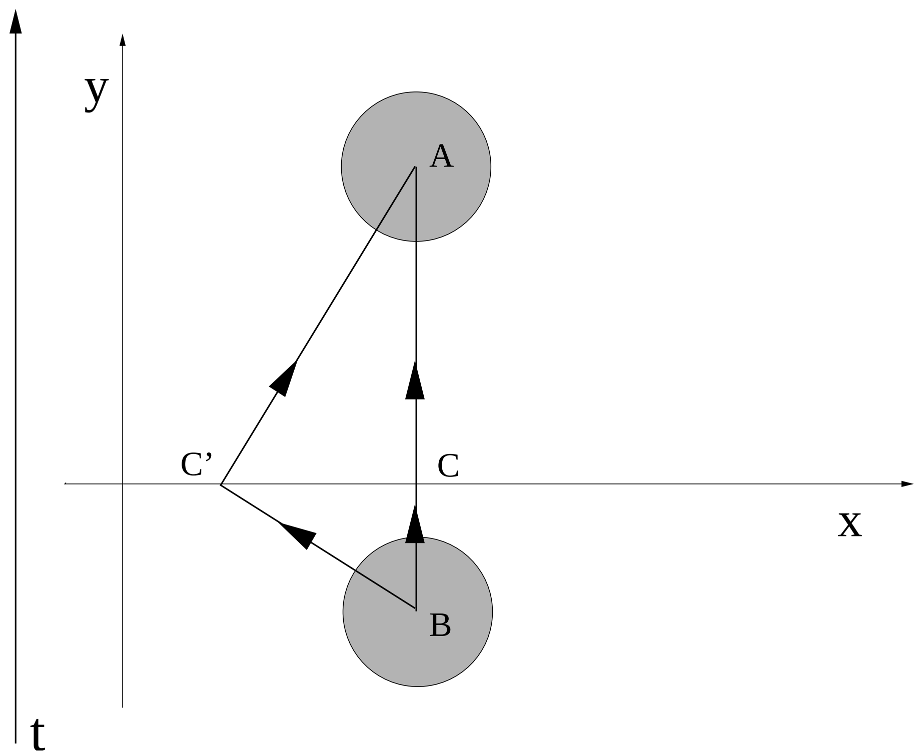


Figure 2.5: Comparing the motion of sphere 2, placed in B, toward sphere 1, placed in A, either through C or through C′.

either along a straight path BCA, or along $BC'A$. In the second case, the path of sphere 2 relative to sphere 1 is not straight, but deviates to a non-minimal distance curve. Stepwise, at any time step, along the non-minimal distance path the relative distance of the two spheres is higher than in the case of minimal distance, and therefore the entropy of the system lower. This configuration is therefore suppressed in the entropy-weighted sum. It is nevertheless still present and accounted for, what precisely makes of this scenario a quantum scenario [15].

So far for what regards the origin of a motion, and its direction. One should wonder where an accelerated motion comes out. Indeed, the configuration we just considered is highly simplified and unphysical. One could ask how is it possible that two spheres, or in general two objects, can be at rest at a certain distance from each other, i.e. how can such a configuration be the one favoured by maximisation of entropy. In fact, one must consider that such a configuration is possible only if one neglects the environment, for instance a device that just placed the two spheres and kept them where they are till the instant before we start our "Gedankenexperiment". Only if we neglect this, we can just start with the two spheres as we have depicted them, and explain how they start moving toward each other, which is the first step of their motion. After the first step, a kinetic energy has been developed. In classical terms, it is potential energy that gets transformed into kinetic energy. In our set up, we just have geometries originating from energy distributions: in this case, we must consider other energy cells to be included in the two-spheres system. This modifies the conditions to new initial conditions for the second step. If we subtract the "kinetic" effect generated at the first step, i.e. if we consider to start once again with the spheres at rest, we can repeat the argument and conclude that once again the system acquires kinetic energy, i.e. motion. This adds to the one already produced at the first step. At least as long as the two spheres are far apart from each other so that the effects of the previous steps can be formally subtracted in a linear way, at any step we add a constant amount of motion. In practice, re-summing up the

[15] In section 3.5 we will discuss how the entropy-weighted sum 2.1.16 can be viewed as a generalization of the Feynman path integral.

effects cumulated at each step, we obtain an accelerated motion.

The set up corresponding to 2.1.16 corresponds to the Einstein's theory only approximately, when one considers the main contribution to the geometry. Once restricted in this way, the equivalence with general relativity is established by the fact that, at large N, the entropic evolution implies a smooth path of minimal steps, that correspond to the minimal-distance motions (i.e. along geodesics) of general relativity: any non-minimal gradient introduced in the motion of geometries corresponds to a smaller symmetry group and therefore smaller entropy, as compared to the smoothest, straightest path. When also less entropic configurations are taken into account, the scenario described by 2.1.16 is no more simply a geometric gravitational scenario, but a quantum scenario. As we will see, it is precisely this feature what implies the introduction of other degrees of freedom besides energy and mass, and other types of interaction besides the gravitational force.

2.4.5 The metric around a black hole

Let us consider once more the general expression relating the evolution of a system as is seen by the system itself, indicated with A', and by an external observer, A, expressions 2.4.1 and 2.4.2. In the large-scale, *classical* limit, the variations of entropy $\Delta S(A)$ and $\Delta S(A')$ can be written in terms of time intervals, as in 2.4.5 and 2.4.6, in which t and t' are respectively the time as measured by the observer, and the proper time of the system A'. As we have seen, in this case expression 2.4.2 can be written as $(\Delta t')^2 = (\Delta t)^2 - \langle \Delta S'_{\text{external}}(t) \rangle$ (see expression 2.4.13), and the temporal part of the metric is given by:

$$g_{00} \; = \; \frac{\langle \Delta S'_{\text{external}}(t) \rangle}{(\Delta t)^2} \, - \, 1 \, . \qquad (2.4.18)$$

As long as we consider systems for which g_{00} is far from its extremal value, expression 2.4.18 constitutes a good approximation of the time component of the metric. However, a black hole does not fall within the domain of this approximation. According to its very (classical) definition, the only part we can probe of a black hole is the surface at the horizon. In the classical limit the metric at this surface vanishes: $g_{00} \to 0$ (an object falling from outside toward the black hole appears to take an infinite time in order to reach the surface). This means,

$$\langle \Delta S_{\text{external}} \rangle \approx \propto (\Delta t)^2 \,. \tag{2.4.19}$$

However, in our set up time is only an average, "large scale" concept, and only in the large-scale, classical limit we can write variations of entropy in terms of progress of a time coordinate as in 2.4.5 and 2.4.6. The fundamental transformation is the one given in expressions 2.4.1, 2.4.2, and the term g_{00} has only to be understood in the sense of:

$$\Delta S(A') \longrightarrow \langle \Delta S(A') \rangle \equiv \Delta t' g_{00} \Delta t' \,. \tag{2.4.20}$$

The apparent vanishing of the metric 2.4.18 is due to the fact that we are subtracting contributions from the first term of the r.h.s. of expression 2.4.2, namely $\Delta S(A')$, and attributing them to the contribution of the environment, the world external to the system of which we consider the proper time, the second term in the r.h.s. of 2.4.2, $\Delta S_{\text{external}}(A)$. Any physical system is given by the superposition of an infinite number of configurations, of which only the most entropic ones (those with the highest weight in the phase space) build up the classical physics, while the more remote ones contribute to what we globally call "quantum effects". Therefore, taking out classical terms from the first term, $\Delta S(A')$, the "proper frame" term, means transforming the system the more and more into a "quantum system". In particular, this means that the mean value of any observable of the system will receive the more and more contribution by less localized, more exotic, configurations, thereby showing an increasing quantum uncertainty. In particular, the system moves toward configurations for which $\Delta x \to\gg 1/\Delta p$. Indeed, one never reaches the condition of vanishing of 2.4.20, because, well before this limit is attained, also the notion itself of space, and time, and three dimensions, localized object, geometry, etc..., are lost. The most remote configurations in general do not describe a universe in a three-dimensional space, and the "energy" distributions are not even interpretable in terms of ordinary observables. At the limit in which we reach the surface of the horizon, the black hole will therefore look like a completely delocalized object.

2.4.6 Natural or real numbers?

The approach we are proposing, and the fact that from the collection of binary codes we arrive to the structures of our physical world, implies a question about what is, after all, the world we experience. We are used to order our observations according to phenomena that take place in what we call space-time. An experiment, or, better, an observation (through an experiment), any perception in itself, basically consist in realizing that something has changed: our "eyes" have been affected by something, that we call "light", which has changed their configuration (molecular, atomic configuration). This light may carry information about changes in our environment, that we refer either to gravitational phenomena, or to electromagnetic ones, and so on... In order to explain them we introduce energies, momenta, "forces", i.e. interactions, and therefore we speak in terms of masses, couplings etc... However, all in all, what all these concepts refer to is a change in the "geometry" of our environment, a change that "propagates" to us, and eventually results in a change in our brain, the "observer". But what is after all geometry, other than a way of saying that, by moving along a path in space, we will encounter or not some modifications? Assigning a "geometry" is a way of parametrizing modifications. Is it possible then to invert the logical ordering from reality to its description? Namely, can we argue that what we interpret as energy, or geometry, is simply a code of information? [16] Something happens, i.e. time passes, when some code changes. Viewed in this way, it is not a matter of mapping physical degrees of freedom into a language of abstract codes, but the other way around, namely: perhaps the deepest reality *is* "information", that we arrange in terms of geometries, energies, particles, fields, and interactions. When we "see" the universe, we *interpret* the codes in terms of maps, from a space of "energies"

[16] To this regard, see for instance the "it from bit" of J. A. Wheeler, and the work of C. F. Weizsäcker.

to a target space, that take the "shape" of what we observe as the physical reality. From this point of view, information is not just something that transmits knowledge about what exists, but is itself the essence of what exists, and the rationale of the universe is precisely that it ultimately is the whole of rationale. The quantum (in the sense of indeterministic) nature of the universe is then the consequence of being any observable not just a code but a collection, a superposition, of codes.

Reducing everything to a collection of binary codes means reducing everything to a discrete description in terms of natural numbers, i.e. to saying that the whole of rationale is numerable. One may wonder whether natural numbers are enough to encode *all* the information of the universe. At first sight, one would say that real numbers say "more", allow to express more information. Moreover, they appear to be "real" in the true sense of something existing in nature. For instance, one can think to draw with the pencil a circle and a diameter. Then, one has *physically* realized two lines whose lengths don't stay in a ratio expressible as a rational number. However here the point is: what is really about the microscopical nature of these two drawings? At the microscopical level, at the scale of the Planck length, the notion of space itself is so fuzzy to be practically lost. In our scenario, an analysis of the superposition of configurations tells us that, before reaching this scale, remote configurations, whose contribution is usually collected under the Heisenberg's uncertainty, count more and more. In other words, the world is no more classical but deeply quantum mechanical, to the point that the uncertainty in the length of the two lines doesn't allow us to know whether their ratio is a real or a rational number. In this sense, this analysis provides further support to an old idea which goes back to Konrad Zuse, that all the information of the universe is expressible through natural numbers, and, as a consequence, the discrete description of the universe, and in particular of space-time, is not just an approximation, but indeed the most fundamental one can think about.

Thinking of the discrete as the ground of a fundamental description of the world makes sense, because in mathematics real numbers are introduced through definitions and procedures, whose informational content can be "written" as a text with a computer program. *As a matter of pure information content*, real numbers are introduced via natural numbers. In our scenario, the fact of summing over an infinite number of configurations as a matter of fact somehow reintroduces in the game the continuum, in a way conceptually reminiscent of the way real numbers are introduced in mathematics. Owing to these aspects, our approach deeply differs from pure "discrete" descriptions of physics. In our case, although not being fundamental but an effective approximation, the continuum is not a concept belonging just to the large scale, but a "built-in" asymptotic limit of the theory.

3 The superstring representation of the universe of codes

3.1 From combinatorics to strings

As discussed in chapter 2, the dominant geometry of the universe at energy N is the 3-sphere of radius $\propto N$. Here the unit of measure can be identified with the Planck scale. In the limit of large N, this scenario can be approximated by a description in terms of interacting quantum particles and fields. Since we start from a description of every observable in terms of geometric distribution of energy, these particles and fields will not simply move inside a space within a well defined geometry, but will determine themselves the geometry. We will have therefore a parametrization of the staple of geometries through propagating fields. In order to have this, we need to associate a fiber to any point (i.e. to any elementary cell of Planck size) of a base. According to the analysis of section 2.2, dimensions other than three are already taken into account by the fact of working with quantum objects. Therefore, the base of the fiber will correspond to the three dimensional space. Let us now come to the field content we must expect to find. We have seen there that the universe behaves like a black hole with horizon at radius $\mathcal{T}$ (where $\mathcal{T}$ is the continuum limit of N), plus "quantum corrections". The horizon expands at the speed of light. It is therefore reasonable to think that it is stirred, or driven, by the propagation of massless fields, such as the photon. To summarise, we must expect that this scenario can be parametrized in terms of a theory containing massless fields, and whose space is given as a fiber over a three-dimensional extended space.

The existence of a minimal length leads to a parametrization in terms of extended objects, i.e. to string theory. By this we mean not just perturbative string theory, but the whole, underlying theory, which includes not only strings but also more in general membranes. Owing to the absolute generality of the combinatorial scenario described in chapter 2, and assuming uniqueness of string theory (or M-theory, or whatever name one wants to give to the theory underlying perturbative string theory), we make here the *hypothesis* that *string theory represents it in the continuum.* Perturbative constructions of string theory give therefore us insight into the theory in terms of fields, elementary particles, and their interactions. However, in order for this equivalence to work, also string theory must be endowed with an entropic mechanism, corresponding to the one at work in the combinatorial scenario. Namely,

i) the string target space must be considered to be compact (i.e. *all* the string coordinates are compactified),

ii) there is no selection mechanism for a specific geometry of compactification, other than a simple stapling of *all* compactification geometries, weighted by their entropy. The latter is related to the amount of symmetry of the "string vacuum" in a way analogous to the relation in the discrete construction of chapter 2. This implies that string entropy must be defined in relation to the volume of the symmetry group of the construction.

Of course, the correspondence between the two scenarios is not a one-to-one correspondence of geometries, because, to start with, as we have discussed in chapter 2, quantization in itself "covers" a whole bunch of geometries, something that reflects in the fact that string theory does not live in an arbitrary number of dimensions. However, we will see that also on the string theory side the entropic mechanism turns out to select three space coordinates as the subspace to be identified with the extended space in which we live.

3.1.1 The logarithmic map

This string scenario is in its ground non-perturbative and in a regime of full interaction. In principle, it accounts for the physics of the universe "pointwise", parametrizing at any time the evolution of the universe of 2.1.16. However, in order to "disentangle" the ingredients of this highly non-perturbative picture, and single them out in terms of elementary degrees of freedom and their interactions, namely, in order to obtain the properties (spectrum, masses, interactions) of the elementary particles as free fields, we must somehow decouple the theory. In order to make possible the construction of the graviton field, we work in a flat space, to be viewed as a local approximation of the real space. This condition leads us to start with supersymmetric string theory in ten dimensions, to be compactified on a product of circles, and then progressively singularized. Even after the "internal" space is twisted, and supersymmetry is broken, even fully broken, there still exists at least one massless field, to be identified with the graviton. Such a mapping of the geometric configuration of the universe is precisely what we need, in order to identify the elementary degrees of freedom. The transverse coordinates of the flat space must be considered as the tangent space to the horizon. The massless graviton can then be viewed as the field that (together with the photon, as we will see), propagating at the speed of light, which is also the speed of expansion of the universe, "stirs" the horizon of space, thereby expanding the universe itself. This interpretation may seem strange, because one would think of identifying the space and fields of the string theory with the local space and fields of the physics around the observer. However, as we will discuss in section 4.1.1.4), there is a duality relating horizon and center of the universe, intended as the point in which the observer is located. According to this, the local flat space approximation turns out to be appropriate also in order to investigate the degrees of freedom of the local physics around the observer.

Once gravity is decoupled by building the construction on a flat space, string theory turns out to give us the real world, but as seen from the tangent space. Instead of telling about the "on shell" physics at any point of space-time, it will give us information on the spectrum of free fields/particles, whose interaction builds up the actual physical space geometry. The complete decoupling is attained by compactifying the string on a product of circles (toroidal compactification). This will be our starting point for the analysis of the sequence of symmetry reductions leading to less and less entropic string constructions. Indeed, stapling less and less symmetric compactifications implies that, at the end, the resulting effective spectrum will correspond to the less symmetric configuration of all, given by the intersection of all the projections. Notice that, as it is defined, the effective physical spectrum must not necessarily coincide with the spectrum of any of the single string compactifications [1]. The details of this analysis, and the spectrum resulting from symmetry minimisation, will be discussed in chapter 4. Here we discuss general properties which allow us to correctly identify the relation between the coordinates appearing in the perturbative string construction we will use for our investigation, and the real, physical space-time coordinates. In the toroidal compactification one can eliminate the time and one space coordinate of the target-space, and work in the light-cone-gauge, which describes only the transverse propagating modes of massless fields. The relation between the representation of space coordinates in the toroidal string compactification to the coordinates of the physical, curved space-time is precisely the one we expect when passing from a description in terms of groups to a description in terms of the associated algebras, i.e., logarithmic. The spheric space-time is here mapped into a two torus. In chapter 2 the entropy of the 3-sphere has been found to be:

$$S_{(3)} \;\sim\; N^2 \,, \tag{3.1.1}$$

whereas the entropy of the circle has been found to have a logarithmic dependence on N:

$$S_{(1)} \;\sim\; \ln N \,. \tag{3.1.2}$$

[1] In this scenario it is not a matter of looking for the "right" string compactification, the one which should produce the physical spectrum of elementary particles and fields as we know it, no more than it is a matter of building a single geometry which exactly reproduces the shape of the physical world. All the observables are here defined as mean values, averaged over an infinite staple of geometries.

Passing from the physical, curved space, to a picture based on a toroidal compactification implies therefore a logarithmic transformation of entropy, and therefore of the coordinate N. The mapping of entropy is:

$$S: \; N^2 \; \to \; 2 \ln N \,, \tag{3.1.3}$$

implying the coordinate transformation $N \to \ln N$, or, in the continuum limit, $\mathcal{T} \to \ln \mathcal{T}$. We see that indeed the string construction turns out to be a realization in a *logarithmic* representation of the real, physical coordinates. We introduce therefore here the concept of *logarithmic picture*, defined as a representation of the physical world through a *staple* of string configurations. It corresponds to an effective theory, in which masses and couplings are computed as mean values on a staple of string compactifications, and are related to the physical quantities by a logarithmic map: the physical quantities are obtained by exponentiation of the relations obtained as functions of the string coordinates in the logarithmic picture. For instance, a mass relation of the type $\mathfrak{m} \sim \alpha / \mathfrak{r} + \kappa$ as obtained in the logarithmic picture corresponds to a physical mass of the type $m \sim \kappa \, R^{-\alpha}$, where $\mathfrak{r}$, the space coordinate in the logarithmic picture, is related to the physical space coordinate R through $\mathfrak{r} = \ln R$ (as we will also do on page 85, we use here Fraktur fonts for quantities in the logarithmic picture, in order to distinguish them from their physical counterparts).

It is well known that a consistent string theory can only be constructed by embedding the string in a higher dimensional target space. The number of these dimensions is fixed by the requirements of supersymmetry (basically needed in order to introduce fermions, i.e. in order to implement a relativistic description of space-time) and quantum consistency, and are apparently not related to the dimension (three) of the space we want eventually describe. However, as seen from our point of view, these two things are deeply related: superstring theory is consistent precisely in the right number of dimensions that make of it the theory which implements a description of the universe we are discussing. In our approach, the number of space dimensions of the universe, three, turns out to correspond to the minimal number of non-twisted dimensions string theory can be consistently reduced to upon compactification, once canonical quantization is imposed. They correspond therefore to the only coordinates along which massless fields are free to expand, after the highest degree of projection and singularization of the string space has been applied. The procedure of singularization of the string target space will be investigated in chapter 4. Here we want to see how the initial ingredients of string theory are precisely those required in order to pop out this result. To this regard, we must first consider that the uncertainty principle, as derived within the theoretical framework of the discrete scenario, implies, and is implied by, only one dimensionality of space, with a well defined geometry. In the combinatorial construction of chapter 2 we have seen that we obtain a "classical" $d = 3$ dimensional space, *plus* the Heisenberg's uncertainty. The dimensionality of space becomes $d = 3 + 1$ once we implement the "time" $\mathcal{T} = E_{\rm tot}$ in a time coordinate suitable for a field theory description. On the other hand, had the dominant dimensionality of space been different from three, in the sum of the rests considered in order to derive the uncertainty (section 2.2) the ratio between weight of the classical and weights of quantum configurations would be different, leading to a different expression of the uncertainty. Moreover, the Heisenberg's uncertainty not only is uniquely related to the dimensionality, but also to the geometry of space, because geometries different from the sphere have different entropy, and therefore different weight, leading to a different uncertainty. Therefore, from the point of view of the scenario of 2.1.16, the Heisenberg's uncertainty not only determines dimension and main geometry, but also the spectrum of the theory. Considering now the realization of this scenario on the continuum, let us see how many internal dimensions do we need. We want to describe all the possible perturbations of the geometry of a sphere in three dimensions, as due to fields and particles that propagate in it. Notice that it is not a matter of building up a set of fields *framed* in a certain space, i.e. functions of space-time coordinates: it is a matter of promoting the deformations of the geometry themselves to the role of fields. One may think of a description in terms of vector fields. Once provided with a time coordinate, the set (3-sphere $\times$ time coordinate), which can be considered the $d = 3 + 1$ "background" space, corresponds to vector fields possessing an $SO(3, 1)$ symmetry. However, we must have both bosons and fermions. Fermions are needed because we want a quantum relativistic description of fields. It is relativity what leads to the introduction of spinorial representations of space. This does not mean we need bosons and fermions in equal number, nor even that they must have the same mass (implying supersymmetry of the theory): supersymmetry is not a symmetry (in the sense of an exact symmetry) of the real world. In terms of spinorial representations, $SO(3, 1)$ is locally isomorphic to $SU(2) \times SU(2)$, a group with 3+3 generators, which, once parametrized in terms of bosonic fields, correspond to a space with six bosonic coordinates. One would like to conclude that, in order to have both a vectorial *and* a spinorial representation of the background space with all its perturbations we need therefore the original 3+1 coordinates *plus* 3+3 internal coordinates. With six internal dimensions it seems we are sure that whatever internal configuration can be mapped to a configuration of space-time, allowing for a non-trivial (and complete) mapping between the "fiber" and the "base" space, ensuring thereby a non-degenerate and complete description of all the perturbations. Ten is precisely the dimension of any perturbative quantum superstring. There is however one more coordinate, obtained by the "un-freezing" of the gravitational coupling, the unit scale, which is indeed the coupling of the theory. Perturbatively, this coupling is flattened into a coordinate (it appears explicitly as such in the 11-dimensional supergravity) [2].

Although the critical dimension of superstring theory is obtained through self-consistency considerations that apparently have little to do with our requirements, it is remarkable that the two dimensionalities turn out to correspond. This provides strong support to the idea that string theory is the right candidate for the translation of

[2]If one wants to retain part of the non-perturbative string description, i.e. with a non-trivial Planck length, he is forced to keep a non-trivial part of the coupling even in a perturbative construction. This may lead to some artifacts, that produce the impression, when looking at just a subset of the construction, that the fundamental theory lives in twelve dimensions (see for instance the works on F-theory, first proposed in [27]).

this scenario. The tight relation we have found between Heisenberg's uncertainty and dimensionality of space, together with the absolute generality of the scenario described by 2.1.16, namely the fact that it considers the collection of all possible geometries, seems to imply therefore also the universality of its translation into the continuum in terms of string theory, providing further support to the idea of the existence of a unique string theory underlying all the possible perturbative constructions.

3.1.2 Entropy in the string phase space

On the string space, 2.1.16 can be translated into:

$$\mathcal{Z}_{\mathcal{T}} = \int_{\mathcal{T}} \mathcal{D}\psi \, \mathrm{e}^{S(\psi)} \,, \tag{3.1.4}$$

where ψ indicates now a whole, non-perturbative string construction, and $\mathcal{T}$ is the time parameter. We recall that, in order to reproduce the discrete scenario, the volume of the string target space has always to be considered finite. The time coordinate too must be considered as being compactified. The sum 3.1.4 is therefore performed over all the string compactifications on a space with the time coordinate of size $\mathcal{T}$. Entropy is defined as usual: as the logarithm of the volume of the symmetry group. An absolute evaluation is not necessary, because the only thing which in practice matters is the relative weight. Since all string compactifications can be viewed as obtained from the toroidal one by acting with projections/symmetry reductions, a relative evaluation is all what is needed, something that leaves undetermined an irrelevant additive constant in the exponent of the integrand, or, otherwise, an overall normalization factor in 3.1.4.

Starting from the most symmetric compactification, perturbatively realized on a product of tori, and proceeding through a chain of symmetry reductions via projection and twisting, we obtain the most singular compactification as the one in which the initial symmetry is reduced to the minimal possible one. It turns out that in this construction all the coordinates of the string target space are twisted, except, in the light cone gauge, for two transverse, corresponding to the "front" of an expanding universe with three space dimensions (see chapter 4). Indeed, as the world described by 2.1.16, and correspondingly by 3.1.4, is given by the stapling of different geometries, also the effective physical description, e.g. the spectrum of elementary particles and their interactions, will be produced by a staple of string configurations, in this case close to the minimum of entropy. Measuring, through an experiment, certain properties of fundamental physics, involves in fact a selection, or projection, in any case an operation of filtering to a particular range of energy scales, sizes, etc... For instance, being interested in the interaction of the electron implies that experimentally one neglects large scale gravitational phenomena such as, to start with, the gravitational field on the earth, or, even more, the cosmological constant. Translated in our scenario, this means that looking for the physics of elementary particles implies neglecting the most entropic configurations, which are the ones that most contribute to the ground curvature of the geometry of the universe.

3.1.3 Pulling-back to the physical picture: the scaling of energy

Consider the typical expression of a mass, or an energy, as computed in a string orbifold:

$$E \;=\; \log \mu \;+\; \big(\text{constants and terms depending on the internal space}\big)\,, \quad (3.1.5)$$

where μ is a scale introduced in order to regularize the computation, and corresponds to the size of a compactified space-time. Compare this expression with the mapping of a typical momentum from the physical space to the logarithmic picture:

$$E_0 \;=\; \frac{k}{\mathcal{T}} \;\stackrel{\log}{\longrightarrow}\; \log \mathcal{T} \;+\; \log k\,. \quad (3.1.6)$$

One recognises the first term of the expression in the logarithmic picture, $\log \mathcal{T}$, as the equivalent of the $\log \mu$ term of 3.1.5, while $\log k$ corresponds to the contribution of the internal space. The latter may depend on moduli, or, whenever the entire internal space is twisted, be a constant. Our previous observations about the re-interpretation of string coordinates in the perturbative string construction as logarithms of the physical coordinates reveal here their importance. What we learn from this comparison is that a quantity of order one in a perturbative string construction does not correspond to a physical quantity of order one (i.e. of order of the Planck scale): as one can see from 3.1.6, once pulled-back to the physical picture, additive constant terms become multiplicative factors, whereas the physical quantity acquires a dependence on the scale of space-time ($\sim 1/\mathcal{T}$) typical of a density. Consider now the energy density of the universe, and the cosmological constant, in a string construction in which the internal space is completely twisted, and supersymmetry is broken. A string computation would give an energy whose dominant behaviour is constant, of the order to the size of the internal space. In a duality-invariant frame, this can be considered of Planck scale size. As a consequence, supersymmetry seems to be broken at the Planck scale, and also the vacuum energy, i.e. the cosmological term, seems to be of the order to the Planck scale. However, for what we have just learned, constant terms are pulled-back to multiplicative factors, that will become densities, acquiring a dependence on the size of the extended space. The energy one computes in a string construction must therefore not be considered a density, but a global value. This represents a deep change of point of view toward the way of approaching string computations. Let us see in detail the transformation from string quantities to the corresponding physical quantities, in order to correctly evaluate the dependence on the physical space coordinates they acquire, and correctly fix the normalization of the transformation 3.1.6. Since in our scenario string theory is defined on a compact space, the vertex operators are not to be normalized by the volume of space, i.e. the volume of the group of translations in the four-dimensional space time. There is in fact no more symmetry under translations, and therefore no overcounting along the orbit of this group, a displacement in space or time representing now an evolution of the universe to a different age. As a consequence, one does not compute anymore densities but global quantities that, in order to be correctly inserted in an effective action, must be divided by an appropriate volume factor of the space-time. A quantity of order one, such as the vacuum energy in the case of supersymmetry broken at the Planck scale, must then be divided by the volume of the base, acquiring a factor $1/\mathcal{T}^2$, the right factor to give the correct size of the cosmological term, as well as the energy density, at present time [3]. Considering string theory as *defined* on a *compact space*, and viewing infinitely extended space only as a limiting case of a compact space, entails therefore a *deep change of perspective*, full of consequences for the interpretation of things that we compute in string theory.

[3] The reason why in the traditional approach string computations produce densities, to be compared with the integrand appearing in the effective action, lies in the fact that space-time is assumed to be infinitely extended. In an infinitely extended space-time, there is a "gauge" freedom corresponding to the invariance under space-time translations. In any calculation there is therefore a redundancy, related to the fact that any quantity computed at the point $\vec{x}$ is the same as at the point $\vec{x}+\vec{a}$. In order to get rid of the over-counting due to this symmetry, one normalizes the computations by "fixing the gauge", i.e. dividing by the volume of the orbit of this symmetry $\equiv$ the volume of the space-time itself. Actually, since it is not possible to perform computations with a strictly infinite space-time, multiplying and dividing by infinity being a meaningless operation, the result is normally obtained through a procedure of regularization of the infinity: one works with a space-time of volume V, supposed to be very big but anyway finite, and then takes the limit $V \to \infty$. In this kind of regularization, the volume of the space of translations is assumed to be V, and it is precisely the fact of dividing by V what at the end tells us that we have computed a density. In any such computation this normalization is implicitly assumed. In our case however, there is never invariance under translations: a translation of a point $\vec{x} \to \vec{x} + \vec{a}$ is not a symmetry, being the boundary of space fixed. On the other hand, a translation of the boundary is an expansion of the volume and corresponds to an evolution of the universe, it is not a symmetry of the present-day effective theory. In our framework, the volume of the group of translations is not V. Simply, this symmetry does not exist at all. There is therefore no over-counting, and what we compute is not a density, but a global value. In our case, compactification of space to a finite volume is not a computational trick as in ordinary regularization of amplitudes, it is a physical condition. In our interpretation of string coordinates, there is therefore no "good" limit $V \to \infty$, if for "∞" one intends the ordinary situation in which there is invariance under translations. In our case, this symmetry appears only strictly at that limit, a point which falls out of the domain of our theory.

3.2 Masses

In the discrete scenario encoded in 2.1.16, massive particles are energy clusters that propagate at a speed lower than the one of expansion of the universe itself, and can therefore be "localized". Like the spectrum, also masses must be explored in the representation in which fields and elementary particles show up, namely, in the string representation. There, masses appear as the lowest momentum of a given particle, and are related to the scale of the universe, which, we recall it, at any time corresponds to a space-time of finite extension. Massive particles and fields correspond to modes that are charged under symmetries of the internal string space. They therefore do not feel just the geometry of the extended, three-dimensional space, but of a higher dimensional space in which some radii are of very small (i.e. string) size [4]. Typical radii are therefore shorter, and, as a consequence, ground momenta higher than just the inverse of the radius of space-time. A varied spectrum of masses is produced by the fact that different particles arise from a staple of string compactifications which have a different amount of projection. They therefore have different symmetry, and also different weight in the phase space. Like energies, in our scenario masses are expected to stay in ratios corresponding to ratios of volumes of symmetry groups, accounting for the weight of the massive state in the phase space:

$$\frac{m_i}{m_j} = \frac{||G_i||}{||G_j||}, \tag{3.2.1}$$

where G_i, G_j indicate the full, non-perturbative symmetry group, whose volume depends on the internal symmetry *and* on the size of space-time. In the case of the elementary particles, these ratios can be analyzed with a good approximation, once we know the staple of string compactifications that mostly contribute to the spectrum of the theory. In the logarithmic picture we obtain mass differences of the

[4] The string length is in principle different from the Planck length. However, in the following and along all this book we will always think in terms of a string-duality invariant reference frame, where the fundamental length coincides with the Planck length.

type:

$$\mathfrak{m}_i - \mathfrak{m}_j \; = \; \beta_j \log \mathcal{T} - \beta_i \log \mathcal{T} \, , \tag{3.2.2}$$

where β_i, β_j correspond to the volumes of the $\mathcal{T}$-independent part of the algebra of the respective symmetry group, and depend only on the properties of the internal string space. In passing to the physical representation, these relations are exponentiated to ratios of the type:

$$\frac{m_i}{m_j} \; = \; \frac{\mathcal{T}^{\beta_j}}{\mathcal{T}^{\beta_i}} \, . \tag{3.2.3}$$

As one can see, heavier masses are not the same as higher momentum excitations. Higher momenta are multiples of a fundamental one, like the higher frequency modes of a string. Different particle masses run instead as *different powers* of the age of the universe.

3.3 Interactions, and couplings

A transition from a particle of higher mass to a (set of) lower mass particles, that is, a decay, always entails a gap of energy, which goes into kinetic energy. This is precisely what, according to our scenario, makes such a transition physically favoured as compared to its non-occurring: it produces a higher spread of energy along space, thereby increasing the symmetry of the geometry, and therefore the overall entropy of the universe. The "coupling" of the interaction depends therefore on the amount of momentum/energy space which is made free by the transition. We define here in all generality the couplings as ratios of weights, i.e. of volumes of symmetry groups. When the symmetry is broken, they can be translated into ratios of masses.

Due to its being the superposition of all possible compactifications, in the universe all symmetries are broken, and this reflects also in the fact that there are no elementary states with the same mass. As we will see, what survives the breaking is the $U(1)$ (gauge) symmetry corresponding to the photon. From a technical point of view, its survival is related to the basic representation of matter as complex fields, a structure explicitly preserved in any superstring construction.

From a physical point of view, the latter are precisely tuned in a way to preserve the spinorial character of the fundamental description of space-time, as required by the combination of quantum mechanics and relativity.

3.4 The strong force

The couplings we have just defined correspond to the weak interaction of the theory, for which an investigation in the perturbative, flat limit constitutes a reasonable approximation. Reintroducing gravitation, and therefore the curvature of space, leads necessarily to the strong coupling of the theory, and to a partial restoration of S-duality. We have seen that particles have masses scaling as different powers of the inverse of the size of space-time. Depending on their interaction, they feel therefore a larger or smaller portion of the "internal" string space. In principle, these masses should correspond to momenta of appropriate subspaces of the whole space. However, we have no recipe for investigating the masses of the modes of the full, non-perturbative, interacting theory. There is however one exception: a compound of particles completely neutral for all the interactions, apart obviously from the gravitational one, should have a classical, or ground, mass corresponding to the inverse of the average radius of the non-perturbative string space. By this we mean the space built up by the staple of geometries, not just the target space of a single string construction, and therefore expressing the result of all the physical interactions and field/matter content. We do not know what is precisely the average geometry resulting from the entropy-weighted staple of non-perturbative string constructions at any time, but we know that it must be some kind of ten dimensional ellipsoid, with three coordinates extended as much as $\mathcal{T}$, the age/radius of the classical universe, and seven internal dimensions of size of order one. If we let a line intersect the ellipsoid by passing through the origin, for any angle of orientation we get a segment, whose length corresponds to the inverse lowest momentum of possible particles, either elementary or not. If we label these momenta with the values of the angle of the intersecting line at the origin, we see that interactions, by gluing together or separating, creating or eating particles, transform angles into other angles. A neutral state is by definition insensitive to interactions, and therefore to variations of the angle. In practice, it feels the space as if it were a symmetric space, a ten dimensional sphere with the same volume (and therefore energy) as the ellipsoid. Its radius is:

$$\widehat{R} \; \propto \; \sqrt[10]{\left(\prod_{i}^{10} R_i = \mathcal{T}^3 \times 1^7 \right)} \; = \; \mathcal{T}^{3/10} \,, \tag{3.4.1}$$

and the corresponding mass:

$$< m > = \; \frac{1}{2} \left(\frac{1}{\mathcal{T}} \right)^{3/10} \, . \tag{3.4.2}$$

This is the mass scale of stable matter, neutral for all the interactions (it is the mass of a hypothetical particle our universe would be made of, had it only gravitational interactions). It corresponds to the typical momentum of a 10-sphere, the most symmetric, and therefore most entropic, geometry with seven internal coordinates of radius 1 and three extended up to radius $\mathcal{T}$. In a temporal average, namely, when events are integrated over time, it results to be the dominant configuration. Since any experiment is performed during an extended time interval, we expect this mass scale to play a fundamental role when comparing with the experimental measurements all the mass evaluations regarding unstable states, i.e. states that exist only for a short interval of time, shorter than the duration of the experimental measurement (see section 4.3.6). As discussed in section 4.3.2.5, it corresponds to the mass of the system $\{$[neutron, proton, electron, neutrino]$\cup$ [their conjugates]$\}$, therefore approximately four times the neutron mass $m_{\rm n}$. This leads to the relation:

$$m_{\rm n} \; = \; \frac{1}{8} \mathcal{T}^{-3/10} \,, \tag{3.4.3}$$

which can be used in order to derive the precise value of the age of the universe to be inserted in the expression of all the other masses and couplings. The neutron mass turns out to be higher than the mass of the bare quarks of lowest mass. This means that the only process of weak decay alone leads to stable matter of weight too low to ensure the existence of a geometric scenario, implying that there must be another type of force at work, stronger than the gravitational one, which counterbalances the electro-weak one. It is the geometry, based on the Planck scale, what requires the existence of both the weak and the strong interaction! At the string level, this is realized through the existence of T-duality. Through this, string theory implements the existence of a minimal length, ensuring thereby that the string is an extended object. Since in the string realization the coupling of the theory too is a coordinate, T-duality results in a so-called S-duality, namely the strong-weak duality.

Much like T-duality, also S-duality is eventually broken. Nevertheless, it does not completely disappear: simply, strongly and weakly coupled sectors are not perfectly symmetrical to each other. A consequence of T- or S-duality is also that there is no perturbative string realization in which all the states and their interactions are visible. The string compactified on circles, as is our case, has momenta and windings, and one cannot wash out the ones or the others: any perturbative realization is based on a choice of limiting procedure, in which one decides which ones have to appear and which of the two (momenta or windings) must be truncated out. In infinite space-time one could think to take a freely-acting orbifold and keep just the ones or the other, thereby realizing perturbatively the full theory. But in this scenario, space is compact, and there is always a part of the theory which is simply "hidden".

3.5 A string path integral

Any string compactification ψ contributing to 3.1.4 describes in itself a "universe" which, along the set of values of $\mathcal{T}$, undergoes a pressureless expansion. In this case, the first law of thermodynamics:

$$dQ \;=\; dU \,+\, PdV\,, \tag{3.5.1}$$

specializes to:

$$dQ \;=\; dU\,. \tag{3.5.2}$$

Plugged into the second law:

$$dS \;=\; \frac{dQ}{T}\,, \tag{3.5.3}$$

it gives:

$$dS \;=\; \frac{dU}{T}\,. \tag{3.5.4}$$

Here T is the temperature of the universe, defined as the ratio of its entropy to its energy. In its dominant configuration, the universe behaves, from a classical point of view, as an expanding, three-dimensional Schwarzschild black hole, and the temperature is proportional to the inverse of its total energy, or equivalently, its radius: $T \;=\; \hbar c^3/8\pi GMk$, where k is the Boltzmann constant and M the mass of the universe, proportional to its age according to $2GM \;=\; \mathcal{T}$. By substituting entropy with energy and temperature in 3.1.4 according to 3.5.4, we obtain:

$$\mathcal{Z} \;\equiv\; \int \mathcal{D}\psi \;\mathrm{e}^{\int \frac{dU}{T}}\,, \tag{3.5.5}$$

where $U \;\equiv\; U(\psi(T))$. If we write the energy in terms of the integral of a space density, and perform a Wick rotation from the real temperature axis to the imaginary one, in order to properly embed the time coordinate in the space-time metric, we obtain:

$$\mathcal{Z} \;\equiv\; \int \mathcal{D}\psi \;\mathrm{e}^{\mathrm{i}\int d^4x\, E(x)}\,. \tag{3.5.6}$$

Let's now define:

$$\mathcal{S} \equiv \int d^4x \, E(x) \, . \tag{3.5.7}$$

Although it doesn't exactly look like, $\mathcal{S}$ is indeed the Lagrangian action in the usual sense. The density $E(x)$ is here a pure kinetic energy term: $E(x) \equiv E_k$. In the definition of the action, we would like to see subtracted a potential term: $E(x) = E_k - \mathcal{V}$. However, the $\mathcal{V}$ term that normally appears in the usual definition of the action, in this framework is a purely effective term, that accounts for the boundary contribution. Let's better explain this point. Usually, in a quantum action in the Lagrangian formulation, one has an integrand of the type:

$$L = E_k - \mathcal{V} \, , \tag{3.5.8}$$

where E_k, the kinetic term, accounts for the propagation of the (massless) fields, and for their interactions. Were the fields to remain massless, this would be all the story. The reason why we usually need to introduce a potential, the $\mathcal{V}$ term, is that we want to account for masses and the vacuum energy (in other words, the Higgs potential, and the (super)gravity potential). In our scenario, non-vanishing vacuum energy and non-vanishing masses are not produced, as in quantum field theory, through a Higgs mechanism, but arise as momenta of a space of finite extension, acted on by a shift that lifts the zero mode (see chapter 4). When we minimise 3.5.7 through a variation of fields in a finite space-time volume, we get a non-vanishing boundary term due to the non-vanishing of the fields at the horizon of space-time (moreover, we obtain also that energy is not conserved). In a framework in which space-time is considered of infinite extension, as in the traditional field theory, one mimics this term by introducing a potential term $\mathcal{V}$, which has to be introduced and adjusted "ad hoc", with parameters whose origin remains obscure [5].

The passage from the entropy sum over string compactifications to the path integral is not just a matter of mathematical trickery. It

[5]Here we have another way to see why the cosmological constant, accounting for the "vacuum energy" of the universe, as well as the other two contributions to the energy of the universe, correspond to densities ρ_Λ, ρ_m, ρ_r, have present values

involves first of all the *reinterpretation* of amplitudes as probability amplitudes. This is on the other hand implemented in the string construction, where quantization is introduced in canonical way. But, besides this, there is something that may look odd at first sight. In the usual quantum (field) theoretical approach, mean values as computed from the Feynman path integral are in general complex numbers, as implied by the rotation on the complex plane leading to a Minkowskian time, $1/T \to it$. Real (probability) amplitudes are obtained by taking the modulus square. This means that what we obtain from 3.1.4, 3.5.6 is somehow the square of the traditional path integral. This is related to the fact that, in order to build up a representation of the fine details of the shape of space, as implied by the staple of energy distributions, we resort to a *spinorial* representation of space-time. Roughly speaking, spinors are "square roots" of vectors. Indeed, as we will see in chapter 4, masses are here originated by a Z_2 orbifold shift of the string space. This shift gives rise to massive particles by pairing left and right moving spinor modes (spinor mass terms in four dimensions are of the type $m\psi\bar{\psi}$). The Z_2 orbifold projection halves the phase space by coupling two parts, and raises the ground momentum. In terms of the weight in the entropy sum, we have at the exponent a pairing/projection $(S(\psi) + S(\psi))/Z_2$, what makes clear that the amplitudes of 3.1.4 are squares of those of the elementary fields (with "weight" $\exp S$). Had we just a vectorial (bosonic) representation of

---

of the order of the inverse square of the age of the universe $\mathcal{T}$:

$$\rho \sim \frac{1}{\mathcal{T}^2} \,. \qquad (3.5.9)$$

Were these "true" bulk densities, they should scale as the inverse of the space volume, $\sim 1/\mathcal{T}^3$. They instead scale not as volume densities but as surface densities: they are boundary terms, and as such they live on a hypersurface of dimension $d = dim[\text{space-time}] - 1$. The Higgs mechanism of field theory itself can here be considered a way of effectively parametrizing the contribution of the boundary to the effective action in a compact space-time. The Higgs mechanism, needed in ordinary field theory on an extended space-time in order to cure the breaking of gauge invariance introduced by mass terms, is somehow the pull-back to the bulk, in terms of a density, i.e. a "field" depending on the point $\vec{x}$, of a term which, once integrated, should reproduce the global term produced by the existence of a boundary.

space, this would not occur, because vectorial (spin 1) or scalar (spin 0) mass terms are of the type $m^2 A^2$, $m^2 \varphi^2$. That is, a mass pairs with *one* boson.

3.6 Resonances

Resonances are a well known effect occurring in physical systems, both at the macroscopic level, for instance in case of momentum transfer between scattering billiard balls, vibrating strings etc..., and at the microscopic level. Of this type are in fact also the absorption of radiation by an atom, or a peak of cross section in the scattering of particles, when a threshold of production of a real particle in the otherwise virtual intermediate channels is attained. In particular, this last phenomenon is used as signal of the existence of particles/fields in high-energy accelerators. Common to all these phenomena is the energy transfer from a system to another one, when the amount of energy corresponds to a typical emission/absorption band. For what concerns the opening of real channels, the effect is formally parametrized by the (denominator of the) field theory propagator, of the type $\sim 1/(p^2 - m^2)$ where m is the mass of the transferred particle or boson, which has a singularity at $p^2 = m^2$, leading to a sudden increase of the (integrated over the momenta and mediated) amplitude. The propagator, on the other hand, shows up as the inverse of the kinetic term of the Lagrangian. In fact, it is already contained in the principle of minimal action, corresponding to the vanishing of the term $T - V$, which translates here into (Kinetic Energy) $-$ (Rest Energy), and as such can also be seen to directly derive from the field theoretical version of the Feynman path integral. This phenomenon appears therefore to be correctly implemented in the theory, and not simply "introduced ad hoc". However, besides the rather refined technical definitions and implementations, the problem of a deeper understanding of resonance is simply translated in understanding why should the evolution of a system be driven by an action principle. In our framework, the entire dynamics is of entropic type, and phenomena do occur simply because they dominate from a simple combinatorial point of view the phase space of all the possible configurations. Entropic are not only all forces, but, as we have discussed, the very existence of a three dimensional universe, and its quantum and relativistic nature.

In our theoretical framework, a resonance occurs whenever the initial energy equals the energy of a state of the theory, and therefore it corresponds to an enhancement in the phase space. In the space of the configurations of energy distributions there is no distinction between "types" of energy: there is only a staple of ways of assigning a certain amount of energy with a certain space distribution. Localizing an amount of energy corresponding to the mass of a particle is absolutely equivalent to producing a particle with the same degree of localization, for the simple reason that the concepts of particle or wave or whatever else belong more to our way of organizing the description of physical phenomena than to the intrinsic essence of physical phenomena in themselves. In this sense, also processes of energy emission and/or absorption in atomic systems are types of resonances, and the smearing of the peak (for instance of absorption) has basically the same origin as the quantum nature of physics itself, namely the fact of being the universe a superposition of geometries. In section 3.5 we have discussed how working with a space-time of finite extension effectively introduces a boundary term that mimics the existence of a rest energy E_0. One can see that E_0 has precisely the right sign to produce in the effective action a kinetic term of type $E - E_0$: an effective action on a compact, truncated space with energy term E is equivalent to an effective action with a lower energy term, $E - E_0$, integrated over the full, infinitely-extended space. Therefore, the entropic approach correctly reproduces the usual kinetic-minus-rest energy term of the effective action that, once inverted, gives the singular term of the propagator, leading to resonance.

An example is the case of the emission of radiation from transitions between atomic energy levels, which has an exponential width, usually formalized in the assumption that a physical photon is a wave-packet of solitonic type, therefore a function of hyperbolic sinus type, i.e. with a Gaussian dependence on the energy spread. In our framework, the Gaussian suppression out of the resonance peak is due to the fact that, being the portion of the universe corresponding to the experiment is a kind of small universe in itself, with total energy $E \sim N$, the geometries corresponding to a different total energy $n < N$ are suppressed by a factor $\mathrm{e}^{n^2 - N^2}$, as if they did correspond to a universe of lower age $n \sim \mathcal{T}' < \mathcal{T} \sim N$ (see the discussion in chapter 2 about the weight of configurations at previous age/lower energy). The Gaussian shape is therefore a consequence of the exponential dependence of the weights on the square of energy [6].

3.6.1 Strong electromagnetic coupling resonances

A fundamental, and key difference, between the scenario we obtain in this theoretical framework, and the traditional approach to quantum field and string theory, is that here, owing to the compactness of space-time, and also of the internal string space, T-duality, and therefore also S-duality, are not completely washed out from the effective physical world resulting from the staple of all the string compactifications. Had we not decoupled the theory, and therefore factored-out the extended space, the geometry of space-time resulting from the staple of string compactifications would automatically account for the presence in certain regions of space-time of physical aspects due to both the dual phases. It is just due to technical reasons that we can only obtain a hierarchy of decoupled constructions with different symmetry, and therefore entropy, giving the impression that S-dual contributions are always suppressed. But this is not the case of the real, physical, interacting theory. When we look at certain specific experimental conditions (e.g. the scattering of a certain type of particles, occurring at a certain center-of-mass energy, etc...) implicitly we have performed a very specific selection of the subspace of the phase-space, in

[6] From this perspective, one could view the inverse-square-power behaviour of the propagator, $1/(p^2 - m^2) = 1/x^2$, itself as the approximation of an exponential (Gaussian) behaviour:

$$\mathrm{e}^{-x^2} \sim 1 + \frac{1}{x^2} + \dots , \tag{3.6.1}$$

and consider thereby ordinary field theory propagation as an approximation of the dynamics of this, more general, scenario.

which S-duality may give detectable contributions.

An example of this situation is provided by certain resonances of the amplitude in the proton-antiproton scattering performed at LHC, that ones tries to explain according to quantum field theory as due to the production of an intermediate Higgs boson (see section 4.5.2 for a discussion and references). This explanation is however for certain aspects controversial, because the signature of these scatterings does not exactly fit with what one would expect from a Higgs production. Interpreting the results as signals of a Higgs boson requires an amount of "model fitting" and adjustments, which otherwise appear as unnecessary within our theoretical approach, where these resonances are explained in a completely different way. Indeed, they are not only justified, but accounted for and predicted, both at the qualitative and quantitative level (see section 4.5.2).

In order to get a rough idea of what is going on, let us consider the electric force between two charged particles of elementary integer charge e. Classically, the electromagnetic energy of the two-particle system is:

$$E_V \sim \frac{e^2}{R^2} \sim \frac{\alpha}{R^2}, \tag{3.6.2}$$

where for simplicity we have neglected all numerical factors and fundamental constants (which can be considered to be set to one). Let us suppose we form a bound state by letting the distance R go "to zero", that is, in our physical framework, to the Planck length: $R \to 1$. For such a state the electric potential energy is simply:

$$E_V \sim \alpha. \tag{3.6.3}$$

The total energy in the rest frame of this state is:

$$E_{p_1+p_2} \sim m_1 + m_2 + \alpha, \tag{3.6.4}$$

where m_1 and m_2 are the masses of the two particles, that we indicate as p_1 and p_2. Let us suppose these two particles are going to produce a strongly coupled bound state. Namely, let us consider the

S-dual situation $\alpha \to \alpha^{-1}$. Having learned that working on a perturbative picture implies working "on the tangent space" of the real physics world, we may expect that, what perturbatively are sums of energies (related to entropies), in a non-perturbative situation should better translate into products of weights in the phase space. Roughly speaking, we should expect a relation of the type:

$$\frac{m_1 + m_2}{m_{p_1 p_2}} \; = \; \alpha \, , \tag{3.6.5}$$

between the total rest energy of the two interacting particles and the mass of their bound state. In the case of the proton-antiproton scattering, at a center-of-mass energy higher than the proton mass by a factor $\alpha_{\rm em}^{-1}$ (the S-dual of the electromagnetic coupling) one can form (pe), $(p\mu)$ intermediate bound states, that enhance the decay channels and therefore the cross section, appearing in the form of a wide resonance of the scattering amplitude around 125 GeV. We will discuss in detail these aspects in section 4.5.

4 The spectrum of the universe of codes

4.1 The non-perturbative solution

The integral 3.1.4 contains in principle all the information about our universe. As discussed in chapter 3, although on the large scale physics is dominated by the configurations of highest entropy, the details of a fundamental description of the microscopic world in terms of elementary particles and their interactions are better investigated by looking at a particular selection out of the bunch of geometries. In particular, when considering the theory in the continuum, the spectrum is investigated by looking at the most singular, i.e. less symmetric (and therefore also less entropic) string configurations. The reason is twofold: on one side, only by looking at the intersection of the most singular configuration it is possible to learn about which symmetry, if any, eventually survives; on the other side, physics of elementary particles can only be experimentally investigated in very "singular" experimental devices, corresponding to very selected configurations (geometries) of the universe. The fact of looking at such an experiment therefore already in itself implies a very targeted operation of selection in the phase space of all possible configurations of the universe.

In order to investigate the physical content of the theory we will use a "perturbative" approach. In ordinary quantum field theory one separates the time evolution into a free propagation and an interaction part. The physical configurations are inspected via the conceptual separation of a base of free states, eigenstates of the free Hamiltonian, which are exact solutions of the free theory. As long as the coupling of the interaction is small, the full solution can be considered a small perturbation of the free propagation, and the perturbative approach makes sense. In our case, we have a truly non-perturbative string system, in which even the space-time is mixed up, and in general will not be factorizable into an extended one, "the" space-time as we experience it, and an internal space. Moreover, we can access the whole theory only through "slices", the perturbative (string) constructions, to be treated as the patches, the "projections", which allow to shed light into the "patchwork", the whole theory. Information about the non-perturbative string properties will be obtained through heavy use of string-string duality. To this purpose, one makes use of properties of (extended) supersymmetry. Unfortunately, this implies working in flat limits (decompactification limits) of the string space. In these limits, the vacuum energy expectation value vanishes. Since we are interested in a description of a space-time of finite extension, this is rather unphysical. In string theory one can explicitly break supersymmetry and end up with a non-vanishing ground energy. However, this situation is anyway artificial, in that the very fact of explicitly observing an operation in a perturbative construction implies working in a decompactified space, and therefore tells little-to-nothing about the real, physical situation. Therefore, we will never see a full description of the whole physical content, expressed in a nice, closed form through a compact formula. Moreover, the traditional computational approach to string scattering amplitudes will not tell us much about the real situation of physical processes. Nevertheless, string theory is a necessary passage toward a better knowledge of the physical content. In our approach, masses, couplings, and amplitudes are related to occupation volumes in the phase space. Their derivation and computation must be performed within this context; the field theoretical approach, with its refined technology, including renormalization, and renormalizability, must be treated here as just an approximation, valid (and unavoidable) when restricted to an appropriate range of fluctuation around reference values derived through investigation of the phase space.

Although the historical reasons that led to the introduction and investigation of space-time supersymmetry are in our framework weakened and, in particular, as we will see, low-energy supersymmetry does not play anymore a key role in the stabilization of mass scales and in justifying a small value of the cosmological constant, nevertheless supersymmetry remains of key importance in the investigation of non-perturbative properties of string theory. It allows in fact to identify dual constructions through the behaviour of certain quantities depending on (and made stable by) a class of states belonging to multiplets of representations of extended supersymmetry. Extended supersymmetry proves therefore to be an exceptional tool in investigating the structure of the string constructions, and we will use the comparison of string duals at the extended supersymmetric level in order to understand the symmetries of the lower (super)-symmetric compactification, when approaching the most singular string vacuum. Although not exactly the explicit formula one would dream of, this will prove to be enough for many purposes, because, in order to investigate ratios of volumes in the phase space, what we need is basically to know what are the *operations* we perform on a construction that we keep under control.

Consistently with the fact that we are investigating a flat limit of the geometry, we will follow the process of symmetry reduction through the spectrum of possible string compactifications in the class of orbifolds. Orbifolds are particular string constructions in which the target space is flat everywhere except from some special points, at which the curvature is concentrated. Having full knowledge of the spectrum of the perturbative states at any energy level, we are able to write the partition function, the "one loop partition function", which in principle encodes all the information about the construction; with this it is possible to explicitly perform one-loop computations of scattering amplitudes and threshold corrections, and therefore compare string duals through pure string computations. Z_2 orbifolds are the best suited for our investigation, because they preserve the basic structure of the target space as a product of circles (it becomes a generic product of circles and orbifolded circles, S^1/Z_2) and mod-out the space by the group with the smallest volume among all the orbifold operations. A product of Z_2 twist/shifts allows therefore to achieve a configuration with a smaller surviving symmetry group than those obtained through any other product of orbifold operations. The most singular orbifold will be the one with the highest amount of freely and non-freely acting Z_2 shifts and twists. Fortunately, Z_2 orbifolds are the easiest and therefore the most investigated constructions [1]. Their investigation allows to get an insight into properties which are typical of string theory in itself: most of the investigations performed at other points in the moduli space must in fact rely on geometrical properties of smooth surfaces, and their singularities. Although for some respects rather powerful, these techniques don't allow to capture the presence of states related to non-geometrical singularities, or even fail in general for the simple reason that, owing to T-duality, the full string space simply cannot be reduced to a geometric one [2].

4.1.1 Investigating orbifolds through string-string duality

Our starting point is a maximally supersymmetric string vacuum with flat background given by a product of circles. The constraints of two-dimensional conformal field theory impose that Z_2 orbifold twists must act on groups of four coordinates at once. Perturbatively, in any string construction there is room for a maximum of three such operations, one of which is however redundant, in that it leads, once combined with the other ones, to the re-introduction in the twisted sectors of the states projected out. Therefore, we can say that only a maximum of two independent Z_2 twists act effectively. However, the amount of supersymmetry surviving to these projections, as well as the amount of initial supersymmetry, is different, depending on whether we start with heterotic, type I, or type II strings. This means that in any construction not all the projections acting on the theory are visible. Indeed, one of them is always non-perturbative. The reason is that, by definition, a perturbative construction is an expansion around the zero value of a parameter, the coupling of the theory, which is itself

[1] See for instance refs. [28, 29, 30, 31, 1, 3, 32, 2, 4, 33, 34, 6, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52].

[2] For examples, see for instance ref. [6].

a coordinate in the whole theory. An orbifold operation acting on this coordinate is forcedly non-perturbative [3]. In the following we will often make use of the language of string compactifications to four dimensions, especially for what matters our reference to the moduli of the string orbifolds. This will turn out to be justified "a posteriori": we will see that indeed the final configuration is the one of a string space with all but four coordinates twisted and therefore "frozen". Only four coordinates remain un-twisted and free to expand, while all the others remain stuck at the string/Planck scale. Massless degrees of freedom move along these and expand the horizon of space-time at the speed of light. Although not infinitely extended, this "large" space is what in our scenario corresponds to the ordinary space-time. The language of orbifold constructions in four dimensions is therefore just an approximation, that works particularly well at large times. Only at a second stage we will discuss how and where this picture must be corrected in order to account also for compactness of the space-time coordinates. Although somehow an abuse of language, this approximation allows us to take and use with little changes many things already available in the literature. In particular, for several preliminary results and a rediscussion of the previous literature, the reader is referred to [6].

Let's see what are in practice the steps of decreasing symmetry we encounter when approaching the most singular configuration. Although the order in which we apply freely and non-freely acting orbifold operations will be at the end irrelevant, it is convenient to organise the analysis by considering first non-freely acting operations, i.e. pure twists with orbifold fixed points. Starting from the M-theory configuration with 32 supercharges, we come, by orbifold projection, to 16 supercharges and a gauge group of rank 16. Further orbifolding leads then to 8 supercharges ($\mathcal{N}_4 = 2$) and introduces for the first time non-trivial matter states (hypermultiplets). As we have seen in [6] through an analysis of all the three dual string realizations of this vacuum (type II, type I and heterotic), this orbifold possesses three

[3] A first investigation of a non-perturbative orbifold, which produces the heterotic string, has been carried out in [53, 54].

gauge sectors with maximal gauge group of rank 16 in each. The matter states of interest for us are hypermultiplets in bi-fundamental representations: these are in fact those which at the end will describe leptons and quarks (all the others are eventually projected out). As discussed in [6], in the simplest formulation the theory has 256 such degrees of freedom. The less symmetric configuration is however the one in which, owing to the action of further Z_2 shifts, the rank is reduced to 4 in each of the three sectors. These operations, acting as rank-reducing projections, have been extensively discussed in [2, 4, 6, 55]. The presence of massless matter is in this case still such that the gauge beta functions vanish. In this case, the number of bi-charged matter states is also reduced to $4 \times 4 = 16$. These states are indeed the twisted states associated to the fixed points of the projection that reduces the amount of supersymmetry from 16 to 8 supercharges.

Let's consider the situation as seen from the type II side. We indicate the string coordinates as $\{x_0, \ldots, x_9\}$, and consider $\{x_0, x_9\}$ the two longitudinal degrees freedom of the light-cone gauge. The transverse coordinates are $\{x_1, \ldots, x_8\}$. Here all the projections appear as left-right symmetric. The identification of the degrees of freedom, via string-string duality, on the type I and heterotic side depends much on the role we decide to assign to the coordinates, as we will see in a moment. By convention, we choose the first Z_2 to twist $\{x_5, x_6, x_7, x_8\}$:

$$Z_2^{(1)}: \quad (x_5, x_6, x_7, x_8) \to (-x_5, -x_6, -x_7, -x_8)\,, \tag{4.1.1}$$

and the second Z_2 to twist $\{x_3, x_4, x_5, x_6\}$:

$$Z_2^{(2)}: \quad (x_3, x_4, x_5, x_6) \to (-x_3, -x_4, -x_5, -x_6)\,. \tag{4.1.2}$$

These two projections induce a third one: $Z_2^{(1,2)} \equiv Z_2^{(1)} \times Z_2^{(2)}$, that twists $\{x_3, x_4, x_7, x_8\}$:

$$Z_2^{(1,2)}: \quad (x_3, x_4, x_7, x_8) \to (-x_3, -x_4, -x_7, -x_8)\,. \tag{4.1.3}$$

Altogether, they reduce supersymmetry from $\mathcal{N}_4 = 8$ to $\mathcal{N}_4 = 2$, generating 3 twisted sectors. Depending on whether we consider the type

IIA or IIB construction, the twisted sectors give rise either to matter states (hyper-multiplets) or to gauge bosons (vector-multiplets). As we discussed in ref. [6], a comparison with the heterotic and type I duals shows that the underlying theory must be considered as the union of the two realizations: owing to the lack of a representation of vertex operators at once perturbative for all of them, for technical reasons no one of the constructions is able to explicitly show the full content of this vacuum. The matter (and gauge) content in these sectors is then reduced by six Z_2 shifts acting, two by two, by pairing each of the three twists of above with a shift along one of the two coordinates of the set $\{x_1, \ldots, x_8\}$ which are not twisted. Each shift reduces the number of fixed points of a Z_2 twist by one-half; two shifts reduce therefore the matter states of a twisted sector from 16 to 4. Altogether we have then, besides the $\mathcal{N}_4 = 2$ gravity supermultiplet, three twisted sectors giving rise each one to 4 matter multiplets (and a rank 4 gauge group). On the type I side, these three sectors appear as two perturbative D-brane sectors, D9 and D5, while the third is non-perturbative. On the heterotic side, two sectors are non-perturbative. As it can be seen by investigating duality with the type I and heterotic string, the matter states from the twisted sectors are actually bi-charged (see refs. [56, 57], and [6]), something that cannot be explicitly observed, the charges being entirely non-perturbative from the type II point of view. The moduli $T^{(1)}$, $T^{(2)}$, $T^{(3)}$ of the type II realization, associated respectively to the volume form of each one of the three tori $\{x_3, x_4\}$, $\{x_5, x_6\}$, $\{x_7, x_8\}$, are indeed "coupling moduli", and correspond to the moduli "S", "T", "U" of the theory. On the heterotic side, S is the field whose imaginary part parametrizes the string coupling: $\mathrm{Im}\, S = \mathrm{e}^{-2\phi}$. It is therefore the coupling of the sector that contains the gravity fields. T and U are perturbative moduli, and correspond to the couplings of the two non-perturbative sectors. On the type I side, on the other hand, two of them are non-perturbative, coupling moduli, respectively of the D9 and D5 branes, while only one of them is a perturbative modulus, corresponding to the coupling of a non-perturbative sector (see [45, 56, 58, 59]). Owing to the artifacts of the linearization of the string space provided by the orbifold construction, gravity appears to be on a different footing on each of these three dual constructions.

4.1.1.1 The maximal twist

The configuration just discussed constitutes the last stage of orbifold twists at which we can "easily" follow the pattern of projections on all the three types of string construction. It represents also the maximal degree of Z_2 twisting corresponding to a supersymmetric configuration. As we will see, a further projection necessarily breaks supersymmetry. The vacuum appears supersymmetric only in certain dual phases, such as the perturbative heterotic representation. Non-perturbatively, supersymmetry is on the other hand broken. This means that, when further twisted, the theory is basically no more decompactifiable: perturbative phases represent only approximations, in which part of the theory content and properties are lost, or hidden. This is what usually happens for instance when one pushes to infinity the size of a coordinate acted on by a Z_2 twist. The situation is the one of a "non-compact orbifold".

The further Z_2 twist we are going to consider is also the last that can be applied to this vacuum, which in this way attains its maximal degree of Z_2 twisting. This operation, and the configuration it leads to, appears rather differently, depending on the type of string approach. Let's see it first from the heterotic point of view. So far we are at the $\mathcal{N}_4 = 2$ level. The next step appears as a further reduction to four supercharges (corresponding to $\mathcal{N}_4 = 1$ supersymmetry). Of the previous projections, $Z_2^{(1)}$ and $Z_2^{(2)}$, only one was realized explicitly on the heterotic string, as a twist of four coordinates, say $\{x_5, x_6, x_7, x_8\}$. The further projection, $Z_2^{(3)}$, acts on another set of four coordinates, for instance $\{x_3, x_4, x_7, x_8\}$. In this way we generate a configuration in which the previous situation is replicated three times. When considered alone, the new projection would in fact behave like the previous one, and produce two non-perturbative sectors, with coupling parametrized by the moduli of a two-torus, in this case $\{x_5, x_6\}$: $T^{(5-6)}$, $U^{(5-6)}$. The product $Z_2^{(1)} \times Z_2^{(3)}$ leaves instead untwisted the torus

$\{x_7, x_8\}$ and generates two non-perturbative sectors with coupling parametrized by the moduli $T^{(7-8)}$, $U^{(7-8)}$. Altogether, apart from the projection of states implied by the reduction of supersymmetry, the structure of the $\mathcal{N} = 2$ vacuum gets triplicated.

The symmetry of the action of the additional projection with respect to the previous ones suggests that the basic structure of the configuration, namely its repartition into three sectors, S, T, U, is preserved when passing to the less supersymmetric configuration. On the type I dual realization of this vacuum, besides a D9 branes sector we have now three D5 branes sectors and a replication of the non-perturbative sector into three sectors, whose couplings are parametrized by $U^{(3-4)}$, $U^{(5-6)}$, and $U^{(7-8)}$.

It is not possible to follow this operation on the type II side, where the coupling of the heterotic construction would appear as a perturbative, internal coordinate ($S \leftrightarrow T$ exchange). The twist in the internal perturbative coordinates is already the maximal one at the $\mathcal{N}_2 = 2$ level. In order to proceed further, one could decide to compactify on a circle also the transversal space-time coordinates (which by the way are eventually going to be considered as compact anyway), and trust them as non-perturbative coordinates of the heterotic string. However, in this case the coordinates to be identified with the physical space-time coordinates would be entirely non-perturbative, and it would be the space-time supersymmetry (i.e. the supercurrents) to be non-perturbatively realized. We would loose in any case the possibility of explicitly following the effect of a further reduction of supersymmetry.

A result of the combined action of these projections is that all the fields S^i, T^j and U^k are now twisted. This means that their vacuum expectation value is not anymore running, but fixed at a scale to be identified with the string-string duality-invariant Planck scale. Nevertheless, for convenience here we continue with the generic notation S, T, U used so far, because it allows to better follow the functional structure of the configuration we are investigating. Twisting of the "coupling" moduli indeed suggests that the geometry is no more decompactifiable. A signal is that, after the $Z_2^{(3)}$ projection is applied, the so-called "$\mathcal{N} = 2$ gauge beta-functions" are unavoidably non-vanishing. According to the analysis of ref. [6], this means that there are hidden sectors at the strong coupling [4]. As a consequence, supersymmetry is actually non-perturbatively broken by gaugino condensation.

Since all the string coordinates apart from the space-time are twisted, supersymmetry is broken at the string scale, a scale which, in a string-string duality-invariant framework is eventually identified with the Planck scale. This is therefore the scale at which, at the same time, not only supersymmetry but also the weak-strong duality (S-duality) is broken. A by-product is also that the space acquires a non-vanishing ground energy, as it should be expected in a real, physical situation (more on this later on).

By looking at the structure generated by this last projection, indeed symmetric to the previous ones, we learn that the matter states of this vacuum are three replicas of the chiral fermions of the theory before the supersymmetry-breaking $Z_2^{(3)}$ projection. The gauge sectors appear as partially perturbative on the type I side. However, the type I vacuum, like the heterotic one, corresponds to an unstable phase of the theory: it appears as supersymmetric although it is not. Moreover, inspection of the gauge beta-functions reveals that they are positive. Therefore, although appearing as free states, the states on the D-branes run to the strong coupling and the apparent gauge symmetries are broken by confinement.

Let's **summarize** the situation. The initial theory underwent three twists and now is essentially the following orbifold:

$$Z_2^{(1)} \times Z_2^{(2)} \times Z_2^{(3)} \,. \tag{4.1.4}$$

In terms of supercharges, the supersymmetry breaking pattern is:

$$32 \ \xrightarrow{Z_2^{(1)}} \ 16 \ \xrightarrow{Z_2^{(2)}} \ 8 \ \xrightarrow{Z_2^{(3)}} \ 0 \ \ (4 \text{ only perturbatively})\,. \tag{4.1.5}$$

[4] We refer the reader to the cited work for a detailed discussion of this issue.

The "twisted sector" of the first projection gives rise to a non-trivial, rank 16 gauge group; the twisted sector of the second leads to the "creation" of one matter family, while after the third projection we have a replication by 3 of this family. The rank of each sector is then reduced by Z_2 shifts of the type discussed in refs. [1, 2, 4], two per each complex plane. As a result, each **16** is reduced to **4**.

On the type I side, the states appear in an unstable phase, as free supersymmetric states of a confining gauge theory, while on the heterotic side they appear on the twisted sectors, and their gauge charges are partly non-perturbative, partly perturbative. The perturbative part is realized on the currents. Like the type I realization, also the heterotic vacuum appears to be an unstable phase, before flowing to confinement; both are indeed non-perturbatively singular, non-compact orbifolds. This reflects on the fact that, as also discussed in [6], both on the heterotic and type I side, perturbative and non-perturbative gauge sectors have opposite sign of the beta-function. This signals that, as the visible phase is confining, the hidden one is non-confining. The matter states of the theory consist therefore of a replica into three families of a bi-charged complex state transforming as $\mathbf{4^w} \times \mathbf{4^s}$, where the $\mathbf{4^w}$ belongs to a weakly coupled sector, while the $\mathbf{4^s}$ to a strongly coupled sector of the theory. Indeed, the fact that 1) with the last twist supersymmetry is broken, 2) the internal string space is curved, and 3) the coupling does not correspond anymore to a modulus but is twisted, frozen at a value of order one in (duality-invariant) Planck units, means that the theory in itself is at the strong coupling, and that a perturbative realization is only possible as a projection onto some subsectors. After further symmetry breaking the $\mathbf{4^w}$ will give rise to the weak interactions, while the $\mathbf{4^s}$ to the strong ones.

In this discussion, we did not consider the details of the non-Abelian gauge groups that arise. Indeed, gauge charges are only visible in the heterotic and type I string constructions. However, technical artifacts highly constrain the possible gauge groups. For instance, in the heterotic construction the embedding of the spin connection into the gauge group always singles out an U(2) (or, at the $SU(2)$ extended symmetry point, $SU(2) \times SU(2)$) factor, which is not present on the type I side.

On the other hand, the gauge sector which explicitly appears in the heterotic construction is an artifact, representing an unstable phase ($\mathcal{N}_4 = 1$) of a gauge sector which is indeed non supersymmetric and at confinement. Therefore, we consider also this factorization as an artifact of the linearization implied by the orbifold construction. Working in a duality-invariant frame implies considering just the rank of the symmetry groups. The only physical distinction that will eventually matter will be whether the symmetry is realized perturbatively as non-confining, in which case we will obtain all broken symmetries, or as a confining sector, non-perturbatively realized, of which we just know the rank and the number of matter states transforming in its fundamental representation. It is only by requiring to *interpret* it in terms of gauge groups that will tell us that in our case the only possible choice is an $SU(3)$ group to be identified with the colour symmetry. Indeed, owing to confinement, also this group is broken. In the following, for the sake of simplicity we will assume the point of view of extended symmetry, namely, ignoring the disappearance of the symmetry due to confinement: n matter states transforming in the fundamental representation of a group will be considered as transforming in the $\mathbf{n}$ of $SU(n)$.

4.1.1.2 Origin of four dimensional space-time

The product (4.1.4) represents the maximal number of independent twists the theory can accommodate: a further twist would in fact superpose to the previous ones, and restore in some twisted sector the states projected out. Therefore, further projections are allowed, but no further twists of coordinates. The twists allow us to distinguish between "space-time" and "internal" coordinates. While the first ones (the non-twisted) are free to expand, the twisted ones are "frozen". The reason is that the graviton, and as we will see the photon, propagate along the non-twisted coordinates, and therefore expand the universe by stretching its horizon, allowing us to perceive these coordinates as our "space-time". We get therefore "a posteriori" the justification of our choice to analyze sectors and moduli from the point of view of a compactification to four dimensions.

4.1.1.3 In how many dimensions does non-perturbative String Theory live?

Besides the above mentioned twists/shifts, the only way to further minimize symmetry is to apply further shifts along the non-twisted coordinates. How many are they? From the type II point of view, there are no further, un-twisted coordinates. But we know that they are there, "hidden" as longitudinal coordinates eaten in the light-cone gauge and in the coupling of the theory. Some of these coordinates appear on the heterotic/type I side as *two* transverse coordinates. If we count the total number of twisted coordinates by collecting the information coming from intersecting dual constructions, and the coordinates which are "hidden" in a certain construction and are explicitly realized in a dual construction, we get the impression that the underlying theory possesses 12 coordinates. For instance, on the heterotic side we have a four-dimensional space-time plus six internal, twisted coordinates, and a coupling. On the type II side we see eight twisted coordinates. We would therefore conclude that the two additional twisted coordinates correspond to the coupling of the heterotic dual. On the other hand, no supersymmetric 12-dimensional vacuum seems to exist, at least not in a flat space: the maximal dimension with these properties is 11. This seems therefore to be the number of dimensions in which non-perturbative string theory is natively defined. Let's have a better look at the properties of supersymmetry. As is known, the supersymmetry algebra closes on the momentum operator. When applied to the vacuum, we have:

$$\left\{ Q, \bar{Q} \right\} \approx 2\, M \,. \tag{4.1.6}$$

From a dimensional point of view, a mass can be viewed as the inverse of a length, so that we can also write:

$$< \left\{ Q, \bar{Q} \right\} > \cong \frac{1}{R} \,. \tag{4.1.7}$$

The supersymmetry algebra suggests that the mass on the right hand side of 4.1.6, in all respects an order parameter for the supersymmetry breaking, could be interpreted as the inverse of the length of a coordinate of the theory. This coordinate refers to an extra internal dimension, or, perhaps more appropriately, to a curvature, i.e. a function collecting the contribution of several coordinates, perturbative as well as non-perturbative. We can therefore view the supersymmetric phase as the limit $R \to \infty$ of a theory with generically broken supersymmetry. This decompactification is only possible if the coordinate R is not twisted. Precisely the fact that, in the breaking of $\mathcal{N}_4 = 2$ supersymmetry to $\mathcal{N}_4 = 1$, the dilaton and the other "coupling" fields get twisted, is a signal that a non-vanishing curvature of the string space has been generated. As we discussed in section 4.1.1.1, this means that, even in the case of infinite volume, we are in a situation of non-compact orbifold. In the orbifold language, this is implemented by the fact that, whenever the coupling field is "explicitated" by going to a dual construction, the corresponding perturbative geometric field appears as a volume of a two-dimensional space. This phenomenon can be observed for reduced supersymmetry (for maximal supersymmetry, there is just the type II string construction). Consider for instance the eleventh coordinate of M-theory, that should correspond to the dilaton of the heterotic string. In the type II orbifold constructions (K3 orbifold compactifications), the heterotic coupling corresponds to a two-torus volume. Considering that this two-dimensional space corresponds, from the heterotic point of view, to "extra-coordinates", one would say that, in order to realize all these degrees of freedom, the full underlying theory should be (at least) twelve-dimensional. However, this is only an artifact of the linearization implied by the orbifold construction, and it means that the simple compactification on a circle is not enough, we need an additional coordinate in order to parametrize a curved space in terms of flat coordinates. From the type II dual we learn that supersymmetry is not restored by a simple decompactification: the string space is twisted [5]. Flatness of the string

[5] In some type II/heterotic duality identifications, the heterotic coupling is said to correspond to un-twisted coordinates of the type II string. This however does

space is broken by a “twist” of coordinates that fixes them to the Planck scale. As a consequence, the supersymmetric partners of the low-energy states are boosted above the Planck scale. In a situation of supersymmetry restoration, they should come down to the same mass as the visible world, and space should become “flat”. However, this is only possible when the twist is “unfrozen” and we can take a decompactification limit, such as for instance the M-theory limit. Otherwise, at the decompactification limit the space becomes only locally flat (non-compact orbifold). Let’s collect the informations so far obtained:

1. As soon as the string space is sufficiently twisted, supersymmetry is broken.
2. Equations 4.1.6 and 4.1.7 suggest in this case a non-vanishing curvature of space.
3. In the class of orbifolds, the phenomenon of curving the string space can only be partially and indirectly seen, through the comparison of dual constructions.
4. These constructions are built on a (perturbatively) flat, supersymmetric background: they provide therefore “linearizations” of the string space.
5. The maximal dimension of a supersymmetric theory on a flat background is 11.

All this suggests that, when supersymmetry is broken, we are in the presence of an eleven-dimensional *curved* background. Any, forcedly perturbative, explicit orbifold realization requires for its construction a linearization of the background. Since a 11-dimensional curved space can be embedded in a 12-dimensional flat space, we have the impres-

not change the terms of the problem: in the artifacts of the flattening implied by the orbifold constructions, part of the curvature may be “displaced”, referred to some or some other coordinates. This “rigid” distribution of the twists, basically dictated by the need of recovering a description in terms of supergravity fields referring to the same space-time dimensionality for both the dual constructions, may induce to misleadingly conclusions. The intrinsic twisted nature of the space has to be considered by looking at the string space in its whole (for more details and discussion, see for instance ref. [6]).

sion of an underlying 12-dimensional theory. However, this is only an artifact; in fact, we never see all these 12 flat coordinates at once: we infer their existence only by putting together all the pieces we can explicitly see. But this turns out to be misleading: the linearization is an artifact.

The 12 dimensional background is only fictitious, we need it only in order to describe the theory in terms of flat coordinates. At the perturbative string level, of these coordinates we see only a maximum of 10.

As a matter of fact, we are however in the presence of a maximum of seven "twisted" coordinates, i.e. coordinates along which the degrees of freedom don't propagate, and four un-twisted ones, along which the degrees of freedom can propagate. By comparison of dual string vacua, we can see that there is room to accommodate more "perturbative" Z_2 shifts: through the heterotic and/or type I realization in the light-cone gauge we can explicitly see two more transverse coordinates which are non-twisted, along which we can accommodate further independent shifts, plus two longitudinal ones, along which no shift can act.

4.1.1.4 Shifting the space-time

Let's count the number of degrees of freedom of the matter states. We have three families, that for the moment are absolutely identical: each one contains 4 (massless) chiral fermions with an "internal" multiplicity 4. The number of matter degrees of freedom is therefore the right one in order to build up three families of *massive* doublets of quarks (with multiplicity 3 out of the 4 of the internal symmetry) and leptons (with multiplicity 1 out of the 4 of the internal symmetry). Indeed, the $\{Z_2^{(1)}, Z_2^{(2)}, Z_2^{(1)} \times Z_2^{(2)}\}$ structure can not only be realized through so-called non-freely acting projections (i.e. pure twists) but also by allowing a fully free action of one or two of these projections. This is obtained by associating to the orbifold twist appropriate shifts along some of the coordinates which are not twisted (see ref. [6], and also [1, 4], for more details). Let us indicate the structure of the "pure-twist" orbifold as (t, t, t). By an appropriate choice of shifts associated to the twists, it is then possible to realize the structures (s,t,t), (s,s,t) and (s,s,s), where s and t respectively indicate the nature of the projection ($s =$ all states shifted; $t =$ pure twist) on the first, second and third complex orbifold plane. The difference between these configurations is that in the (t,t,t) realization we have a replication of the matter states into three families, in the (s,t,t) realization we have just two families, in the (s,s,t) one family, whereas in the (s,s,s) there is no matter at all. All these constructions belong to the string phase space, and contribute to the overall appearance of the string realization of the scenario described by 2.1.16. The existence of three different realizations of the $Z_2 \times Z_2 \times Z_2$ orbifold plays a key role for the mass differentiation between matter families: owing to this the most singular configuration results from a superposition in which one family is present only one time, one family two, and one three. In the logarithmic picture the ratio of their phase space occupation volumes is therefore 3:2:1. However, as long as only operations acting on the "internal" string space are concerned, all matter states are massless. Masses are introduced by shifts acting on the non-twisted coordinates. These are the coordinates to be identified with the space-part of space-time. Matter states "projected out" by this kind of orbifold operation are not thrown out from the spectrum of the low energy theory, but acquire a mass related to the scale of space-time. Compactifying the coordinates of space-time implies a change of perspective as compared to the usual approach to string theory. The bosons X^{μ}, X^{ν} are no more to be considered as "living in a space framework" of infinite extension, but describe an expanding space of finite volume. Massless fields such as the graviton (and the photon, of which we will talk in the next section) expand the space by stirring the horizon. The light-cone gauge can be considered as parametrizing the tangent space to a point of the horizon, as illustrated in picture 4.1. Normally, a horizon is a curved surface that works as the boundary of a flat space. For instance, a 2-sphere as boundary of a 3-ball. Nevertheless, the horizon of our physical universe encloses a region of curved space. The reason is due to an artifact produced by the propagation of light. The 2-sphere that seems to have volume (surface) $\pi \mathcal{T}^2$, where $\mathcal{T}$ is the age of the

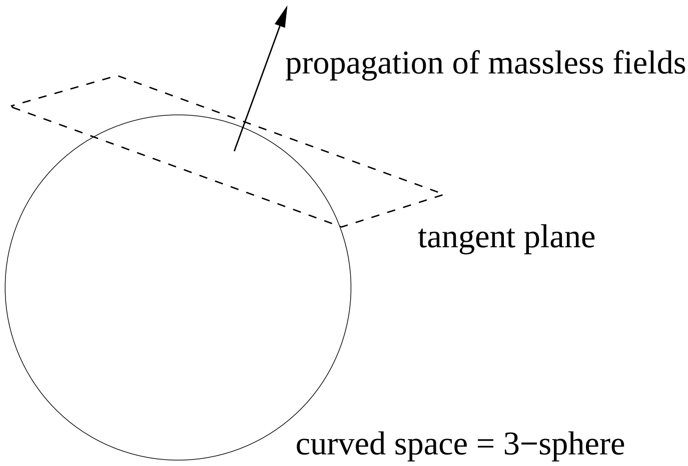


Figure 4.1: The physical space is here represented by a ball. It is indeed a 3-sphere (plus quantum corrections). The transverse coordinates of the perturbative string construction represent the tangent space, out of which the graviton propagates along the normal to the tangent plane, stirring the horizon and thereby expanding the universe.

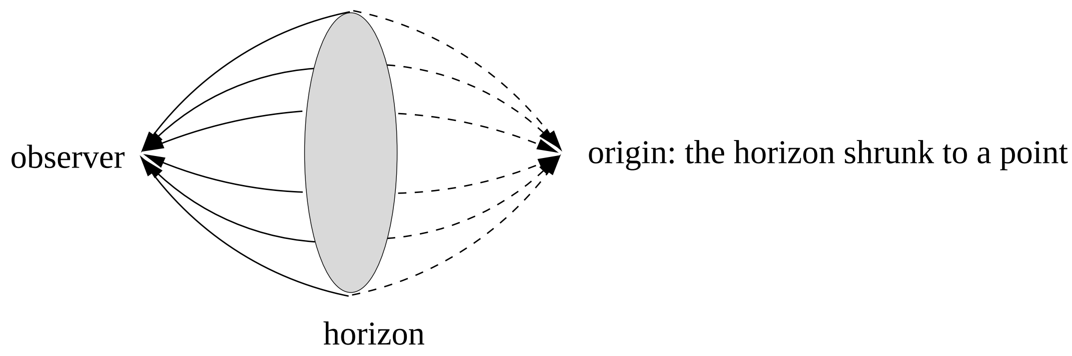


Figure 4.2: The horizon of the universe, here sketched in a simplified way as a disc, in reality a 2-sphere, shows the origin (in temporal sense) of the universe. In our scenario, where the universe expands at the speed of light, this implies it is also the origin in spatial sense. It corresponds therefore to a point (of Planck size). When shrunk to a point, the space "flowing" from the observer to the point at the origin constitutes a curved space, with curvature $\sim 1/\mathcal{T}^2$. This is the curvature of a space with energy content corresponding the actual value of the cosmological constant Λ.

universe as expressed in light years, indeed corresponds to a "point", the origin of the universe. Whereas in general this point is to be intended just in atemporal sense, in our scenario, where the universe expands at the speed of light, this is a point (of Planck size) also in spatial sense. It is his dual interpretation/identification what allows to see the space as curved (see figure 4.2) to a geometry consistent with the value of the cosmological constant.

Let us see what are the possible operations that lead to the most singular compactifications. Along the two transverse coordinates of space time, which, although of large extension, are anyway compact, it is possible to act with two independent Z_2 shifts. Each of them may in turn act either on the momenta, or on the windings of the bosonic lattice (or even on both at the same time). In all these cases, there are states which acquire a mass. Depending on the kind of action, the latter can run as the inverse of the radius of compactification: $m \sim 1/R$ (momentum-shift), or as the radius: $m \sim R$ (winding-shift), or in a T-duality-invariant way, $m \sim 1/R + R$ (momentum- and winding-shift). We obtain therefore a whole bunch of situations, that staple up to produce the spectrum of elementary particles and fields, that we now analyze in detail. The pairing of the shift with the orbifold twist can be of two kinds: either 1) it acts by lifting the mass of *all* the states of a twisted sector, or 2) it acts on linear combinations of the states, in such a way to reduce the rank of the symmetry group, or the number of states, by a factor two. Both these kinds of operation are considered in ref. [6].

Let us consider the first type of pairing, case 1). In this case, all the states twisted by the corresponding Z_2 twist become massive. Depending of whether the shift acts on the momenta or on the windings, or on both, we obtain an over-Planckian mass (winding-shift) or a sub-Planckian mass (pure momentum-shift, $m \sim 1/R$). In the first case, all the states disappear from the low-energy spectrum. In the second case, they become massive states of the low-energy spectrum, with a sub-Planckian mass. They are therefore experimentally observable. The four chiral fermions of a twisted sector must now be interpreted as two massive fermions. As a consequence, the $SU(4)$ acting on chiral fermions becomes an $SU(2)$ acting on massive states. Indeed, it is a broken $SU(2)$, because the shift lifts also the mass of the gauge bosons. Strictly speaking, there is no more gauge group, but, in case of sub-Planckian mass, it is still possible to speak of massive bosons. However, the operation on the bosons cannot be explicitly observed, being these states non-perturbative in a construction (the heterotic one) in which the matter states are visible, and realized on the twisted sector.

In case 2), only half of the matter states is mass-lifted. We are left with two massless chiral fermions transforming under one of the two $SU(2)$ subgroups of $SU(4) \supset SU(2) \times SU(2)$. In case of momentum-shift, when the stapling of this configuration with the one of case 1) is considered, the situation must be interpreted as describing a parity-breaking interaction of two massive fermions in which only one of the two chiralities transforms under an $SU(2)$ symmetry. We can call the two chiral spinors $\psi_{\mathrm{L}}^{(1)}$, $\psi_{\mathrm{L}}^{(2)}$, and the surviving symmetry group $SU(2)_{\mathrm{L}}$. From the *staple* of the two configurations we obtain therefore the realization of the parity-breaking chiral (i.e. only left-handed) coupling of the weak interaction, realized as a "lightly" broken symmetry with bosons of sub-Planckian mass.

In case of winding- (or momentum+winding-) shift, we do not have anymore in the spectrum the right-handed part of the matter states. We just have two massless chiral spinors. Owing to the absence of chiral partners, and, still owing to the fact that the configuration will eventually staple with the previous one, by consistency we are left with the only possible interpretation of these degrees of freedom as the left-and-right moving part of a single particle, which will eventually become massive thanks to the stapling, but that must be considered as a linear combination of the degrees of freedom of the other configurations. Since the $SU(2)$ that rotated just the left-moving part of the massive fermions is now "transversal" to the states of this construction, which are the left- and the right-moving part of a single particle, we cannot consider these two degrees of freedom as rotated into each other by $SU(2) \equiv SU(2)_{\mathrm{L}}$: they are rotated into each other by another symmetry, deriving from the breaking of an $SU(2)$ transversal to $SU(2)_{\mathrm{L}}$, however still a subgroup of the initial $SU(4)$. It is an $SO(2)$ symmetry, that we interpret as $U(1)$. The symmetry of the construction tells us that also the other two degrees of freedom possess a similar symmetry, that we cannot see just because we cannot view gauge symmetries once the states are massive. However, in this way, we can see where the mass gap between pairs of the broken $SU(2)_{\mathrm{L}}$ symmetry group comes from. The stapling of all these operations realizes therefore the breaking of $SU(4)$ to $U(1) \times SU(2)_{\mathrm{L}}$. The massive state whose left and right moving part correspond to these two degrees of freedom has therefore the following transformation properties under $U(1) \times SU(2)_{\mathrm{L}}$: its left-moving part is charged under $SU(2)_{\mathrm{L}}$, while the right moving part is uncharged. Both the left and right moving part are charged under $U(1)$. Since it derives from the Cartan of an

$SU(N)$ group, this $U(1)$ is necessarily traceless. As we will see in section 4.1.2, the tracelessness condition is not realized simply among the degrees of freedom deriving from the initial **4** of $SU(4)$: the introduction of a mass gap implies also different charge, i.e. interaction, properties, leading to a different weight in the phase space. Each one of these degrees of freedom transforms also under a non-perturbatively realized representation of a **4** of $SU(4)$. The condition of tracelessness is realized by summing on this "internal" index. In section 4.1.2 we will come back to the point that the orbifold twist that reduces supersymmetry from $\mathcal{N}_4 = 2$ to $\mathcal{N}_4 = 1$ indeed breaks supersymmetry completely to $\mathcal{N}_4 = 0$ while sending part of this internal sector to the strong coupling. We will discuss how this implies the breaking of the **4** into $\mathbf{1} + \mathbf{3}$, giving rise to the separation in leptons and $SU(3)$ quark triplets.

In the orbifold construction, based on a factorization of space, the space part of space-time is realized as a product of circles, two of which appear as independent transverse coordinates. It would seem that independent orbifold operations (e.g. a combined action of both the two above described shifts) are allowed. However, although possible, and therefore present in the sum 2.1.16, configurations leading to an asymmetrical ground geometry of space are entropically unfavoured: entropy favours a geometry in which a massless field such as the graviton expands the universe, by stirring its horizon, in a symmetrical way, i.e. producing the geometry of a sphere. Of course, the stapling geometries, and the presence of matter states associated to shifts along the space coordinates, will eventually break this symmetry. However, this is a second order effect, a "soft" breaking. For the purpose of the present discussion, we must assume that the independence of the two transverse space coordinates is an artifact of the linearization of space introduced by the factorization of the string space into a product of circles. As soon as the string space is curved, consistently with a non-vanishing net matter/energy content, the extended space is more like a sphere (see chapter 2), with only one radius, and radial coordinates. It does not make sense to consider combinations of the two shifts above described, as they would seem to be allowed by picking independent shifts along the two circles of the transverse extended space, nor to distinguish whether a shift is taken along one or both of these coordinates. The configurations we have described, obtained by applying only one of the two shifts at once, exhaust all the possibilities.

We are now in a position to refine the evaluation of the ratios of volumes introduced by the stapling of configurations with one, two, and three matter sectors. Considering just the spread of configurations produced by shifts acting on the internal coordinates we arrived to the conclusion that these ratios are 3:2:1. Indeed, these states become massive due to shifts along the space coordinates of space-time, and the combinatorial possibilities of realizing these shifts must be taken into account in order to give a finer evaluation of the ratios of volumes. Up to permutations of the three sectors, and the two transverse space coordinates, in the case of just one twisted sector (orbifold "(t, s, s)") we have only one possibility for the momentum shift of type 1). In the "(t, t, s)" orbifold, we have the possibility of mass-shifting one or two sectors. Therefore, when their stapling is considered, one family gets a mass "twice" as large as the other one. In the "(t, t, t)" orbifold, there are always at least two sectors which get mass-shifted: in one case, when the momentum shift involves just one orbifold twist, we have two massive sectors; when it involves two (independent) orbifold twists (and the shift is taken along the other of the two space coordinates), all the three twisted sectors are massive. Let us now consider stapling all these configurations on top of each other. Since families do not bear a label, but are just identified according to their mass properties, we must consider to staple the average volumes, or masses, of any type of orbifold. Normalized to the (t, s, s) case, which therefore has conventionally volume 1, the shifted (t, t, s) orbifolds contribute to an average mass $(2 + 1)/2$ for two families, and the (t, t, t) to an average mass $(2 + 2 + 1)/3$ for three families. The lightest family appears in only one case, namely when there are three twisted sectors ((t, t, t) case), then at a higher level we have the second family, which appears when we have two and three twisted sectors ($(t, t, s) \cup (t, t, t)$ case), and finally the heaviest family, which is the one appearing in all three orbifold cases ($(t, s, s) \cup (t, t, s) \cup (t, t, t)$ case). The mass ratios formerly given as 3 : 2 : 1 are therefore corrected to:

$$V^{(3)} : V^{(2)} : V^{(1)} \;\simeq\; \left[\frac{5}{3}+\frac{3}{2}+1\right] : \left[\frac{5}{3}+\frac{3}{2}\right] : \frac{5}{3}\,. \tag{4.1.8}$$

In order to pass to the concrete computation of masses, first of all the volume ratios 4.1.8 must be transformed into mass ratios. Let us introduce the coefficients $A_{(i)}$, $i = 1, 2, 3$ defined as $A_{(1)} \equiv 5/3$, $A_{(2)} \equiv 5/3+3/2$, and $A_{(3)} = 5/3+3/2+1$. Since they were computed in a logarithmic realization of the physical geometry, in terms of the "real" coordinates they correspond to mass expressions of the type:

$$\ln \mathrm{m}_{(i)} \;\sim\; A_{(i)}\kappa \ln \mathcal{R}\,, \tag{4.1.9}$$

where we have introduced the real mass m, the real radius of space, $\mathcal{R}$, of which the orbifold radius is a logarithm ($1/R \sim \kappa \ln \mathcal{R}$), and allowed for the presence of a coefficient κ, because as yet we did not discuss the overall normalization. Mass ratios between families are therefore of the type:

$$\frac{\mathrm{m}_{(i)}}{\mathrm{m}_{(j)}} \;\sim\; \frac{(\mathcal{R}^{\kappa})^{A_{(i)}}}{(\mathcal{R}^{\kappa})^{A_{(j)}}}\,. \tag{4.1.10}$$

Mass ratios will however be the same if all masses are multiplied by a common factor, which may depend on $\mathcal{R}$. Indeed, considering the orbifolds (t, t, t), (t, t, s) and (t, s, s) tells only about mass differences through generations. However, all particles receive also a mass contribution from the "zero" sector, namely the (s, s, s) orbifold, which does not contain matter states, but can bear a shift along the space coordinates as well, contributing to a ground energy of the matter sector. This is the configuration with no matter, and it contributes with a vanishing term to the sum of contributions to a mass in the logarithmic picture. It must be considered as the scale reference (the zero of the scale) in the additive representation of what, out of the tangent space, are products of weights. The (s, s, s) space-shifted orbifold provides therefore for a multiplicative factor, the factor common to all masses.

In the logarithmic picture, the "vacuum" contribution in the matter sector is:

$$\ln \mathrm{M}_0 \; = \; -\frac{1}{2} \ln \mathcal{R} \, . \qquad (4.1.11)$$

When pulled back to the physical picture, it gives a ground rest energy contribution:

$$\mathrm{M}_0 \; = \; \frac{1}{2\sqrt{\mathcal{R}}} \quad \left(\equiv \frac{1}{2\sqrt{\mathcal{T}}} \right) . \qquad (4.1.12)$$

The factor $\frac{1}{2}$ normalizing the mass is here introduced because R is a radius, and masses are the lowest momentum in a compact space with periodicity given by the full length of space, therefore twice the radius.

The coefficient κ in 4.1.9 and 4.1.10 is calculated by taking into account the amount of projections actually acting on the matter sector in order to break the symmetry not only, as we did, between families, but also within each family, leading to a hierarchy of particles subdivided into leptons and quarks (breaking of S-duality), each of them in turn divided into "up" and "down" of a broken $SU(2)$. A detailed computation will be done in section 4.2, and it will end up in the evaluation of the volume of the broken symmetry factor between the lightest particle of the first and second family, an $SU2)$ symmetry which cannot be considered a gauge symmetry being the gauge bosons completely absent from the sub-Planckian spectrum. This will give us the ratio:

$$\frac{\mathrm{m}_{(2)}}{\mathrm{m}_{(1)}} \; \sim \; \mathcal{R}^{\kappa(A_{(2)} - A_{(1)})} \, , \qquad (4.1.13)$$

as a function of the $SU(2)$ coupling. In this way we will determine the coefficient κ, and from 4.1.8, 4.1.10 and 4.1.12 the absolute mass values of the three lightest particles (to be eventually identified with the neutrinos) given as:

$$m_i \; = \; \mathrm{M}_0 \times \mathrm{m}_{(i)} \quad i = 1, 2, 3 \, . \qquad (4.1.14)$$

Notice that the shifts along the space coordinates break the Lorentz symmetry. Therefore, the superposition of differently shifted configurations not only implies the breaking of parities, but also the breaking of the symmetry of space under rotations. This occurs at the same time as masses are produced: the amount of breaking of space rotations produced is of the same order of the particle masses.

4.1.2 The photon and the $SU(3)$ of QCD

Let us go back for a moment to the $Z_2^{(1)} \times Z_2^{(2)} \times Z_2^{(3)}$ orbifold point, before the introduction of shift operations on the extended space coordinates. Since this orbifold is symmetric under the exchange of each projection, the superposition of configurations which breaks the symmetry in the weak sectors analogously breaks also the strong sector: also the internal $\mathbf{4}$ of each bi-charged fourth-plet gets broken. The pattern of the breaking must therefore be compatible with an effective description in terms of gauge field theory. This requires that, since the gauge sector is at the strong coupling, the *interpretation* we give to this breaking is that the initial $\mathbf{4}$, corresponding to $SU(4)$, has been broken into $\mathbf{1} \oplus \mathbf{3}$, i.e. the gauge group to $U(1) \times SU(3)$. Only in this way we have in fact gauge and matter degrees of freedom in the right amount to give rise to a confining gauge group representation, that we identify with the quarks colour group: $SU(3) \equiv SU(3)_c$. The four states transforming in the $\mathbf{3}$ are to be identified with the quark colours, whereas the singlet is a lepton. Since all these states factorize an $U(1) \times SU(2)_{\mathrm{L}}$ symmetry index, they split into "up" and "down" of the $SU(2)_{\mathrm{L}}$ symmetry. Conversely, one could say that, for each of the three matter families, the doublet of states charged under the chiral $SU(2)_{\mathrm{L}}$ broken symmetry group breaks into a singlet and a triplet of $SU(3)_c$. Since the $\mathbf{3}$ is strongly coupled, the three degrees of freedom are in practice paired, so that the symmetry breaking is effectively a $\mathbf{4} \to \mathbf{1}_1 + \mathbf{1}_2$ breaking, the $\mathbf{1}_1$ being a trivial singlet of $SU(3)$ corresponding to a state charged only under $U(1)$, the $\mathbf{1}_2$ being instead a singlet of $SU(3)$ made out of three charged states.

Once space-time is shifted as described in section 4.1.1.4, the fact of having different symmetry properties produces a mass gap between $\mathbf{1}_1$ and $\mathbf{1}_2$, which adds to the mass gap introduced by the breaking of $SU(2)_{\mathrm{L}}$. Of course, quarks are expected to weight more than leptons, because they bear a further symmetry index.

All these states are charged under $U(1)$. Like the one singled out by the symmetry breaking in the perturbative gauge sector analyzed in section 4.1.1.4, deriving from the breaking of an $SU(N)$ symmetry also this $U(1)$ is traceless. It combines with the other one to give rise to a unique $U(1)$ symmetry, the only symmetry that survives in this symmetry breaking scenario [6]. Coming from the breaking of an $SU(4)(\times SU(4))$ symmetry, the $U(1)$ factor is traceless. This means that it acts by transforming with opposite phase states charged under $SU(3)$ and uncharged ones:

$$\begin{aligned} U(1)\,\varphi &= \mathrm{e}^{\mathrm{i}\beta}\,\varphi\,, \\ U(1)\,\varphi_a &= \mathrm{e}^{-\mathrm{i}\,\beta/3}\,\varphi_a\,, \qquad a\,\in\,\mathbf{3}\,\mathrm{of}\,SU(3)\,. \end{aligned} \tag{4.1.15}$$

Here φ indicates a full chiral fourth-plet of the weak sector. These states, as we have just seen, arrange into massive doublets of a broken weakly coupled $SU(2)$, that we identify with the symmetry group of the weak interactions. The condition on the trace of $U(1)$ holds for $SU(2)$ doublets, but tells nothing about the charge assignments among the states of each $SU(2)$ pair. This indication comes from a further condition, namely the fact that the *strength* of the $U(1)$ interaction is in this context by definition related to the weight this interaction has in the phase space of all the configurations. Since quarks occur three times more than leptons (remember that each fourth-plet, or equivalently each $SU(2)$ doublet pair, bears an internal multiplicity $\mathbf{4} = \mathbf{1}_{\mathrm{leptons}} + \mathbf{3}_{\mathrm{quarks}}$), we obtain the following condition on the charge Q:

$$\sum_{\mathrm{quarks}} |Q(U(1))| \;=\; 3 \sum_{\mathrm{leptons}} |Q(U(1))|\,. \tag{4.1.16}$$

Besides this, we have also the condition 4.1.15 on the trace that in

[6] From a physical point of view, the two $U(1)$ are the same symmetry group: what is split into two sectors is only our representation of this physical situation. Indeed, we have to deal with the very same states, which bear two $SU(N)$ indices ($SU(2)_{\mathrm{L}}$ and $SU(3)_c$), and a $U(1)$ charge.

terms of the charge can be written as:

$$\sum_{\text{leptons}} Q(U(1)) \; = \; - \sum_{\text{quarks}} Q(U(1)) \, . \qquad (4.1.17)$$

The fact that these conditions hold separately for each of the three matter families implies that in each family there must be one state with $Q(U(1)) = 0$. This must necessarily be identified with the lightest particle of each family. If we call the leptons of the fourth-plet in the usual way neutrino and electron, and the quarks down and up quark, and set by convention $Q_e = -1$, from 4.1.15, 4.1.16 and 4.1.17 we derive the charge assignments $Q_\nu = 0$, $Q_u = 2/3$, $Q_d = -1/3$. The $U(1)$ gauge group has all the characteristics of $U(1)_\gamma$, the group of electromagnetism. The corresponding vector field is the photon, and the neutrino, being the less interacting particle, must be identified with the lightest of the fourth-plet.

The spectrum does not contain the degrees of freedom of a possible Higgs boson. On the other hand, here there is no need of such a field, because masses are generated with a pure stringy mechanism, and are basically related to the compactness of the whole space. As remarked in section 3.5, the Higgs boson of ordinary field theory can in some way be thought of as the parametrization of a boundary term through a field propagating in the bulk of space [7] (in section 4.5.2 we

[7]It is legitimate to ask what is the mass scale of the gauge bosons of the "missing" $SU(2)$, the would-be $SU(2)_\mathrm{R}$ of the original weak fourth-plet, $\mathbf{4} = \mathbf{2}_\mathrm{L} + \mathbf{2}_\mathrm{R}$. Namely, asking whether there is a scale at which we should expect to observe an enhancement of symmetry. The answer is: there is no such a scale. The reason is that the scale of these bosons is simply T-dual, with respect to the Planck scale, to that of the masses of particles. Let's consider this shift as seen from the heterotic side. On the heterotic vacuum, matter states originate from the twisted sector, while the gauge bosons (the visible gauge group, the one involved in this operation) originate from the currents, in the untwisted sector of the theory. Similarly, on the type I side, gauge bosons and the charged states we are considering originate from D-branes sectors derived respectively from the untwisted, and the twisted orbifold sectors of the type II theory they were derived from (The type II vacua are on the other hand not appropriate for the investigation of this phenomenon, because the gauge charges are non-perturbative. In any case, although in the form of just the Cartan subgroup of their symmetry group, gauge bosons and matter states arise from mirror constructions, related each other by the type II dual of the heterotic T-duality

will comment about the 125 GeV resonance detected at LHC [60], and usually seen as a signal of the Higgs boson).

4.1.3 The fate of the magnetic monopoles

Under the conditions of the scenario we are discussing, namely of a universe "enclosed" within a finite, compact space, also the issue of the existence of magnetic monopoles changes dramatically. Magnetic monopoles can be of two kinds: the "classical" ones, namely those associated to a non-vanishing "bulk" magnetic charge that parallels the electric charge in a symmetric version of the Maxwell's equations, and the topological ones. In our scenario there are neither classical nor topological monopoles. The existence of classical monopoles would be possible only in the absence of an electromagnetic vector potential, what we have called the "photon" A_μ; their existence has therefore been ruled out as soon as we have discussed the existence and the masslessness of this field. The first idea about the existence of magnetic monopoles in the classical sense (i.e. non-topological) originated by a request of symmetry: were not for the absence of magnetic charges, the Maxwell equations would be completely symmetric in the electric and magnetic field. However, the symmetry of these equations, preserved in empty space, is precisely spoiled by the presence of matter states that are also electrically charged. In our scenario, the description of the universe is "on-shell" and the presence of matter comes out as "built-in": it cannot be disentangled from the existence of space itself. In this scenario there are no topological monopoles either. Since all vector fields are twisted (i.e. massive at the Planck scale or above it) with the only exception of the photon A_μ, propagating in the four-dimensional space time, and since this space-time dimensionality is electro-magnetically self-dual, the only possible topological

under consideration, see discussion in ref. [6]). It is therefore clear that a shift on the string lattice lifts the masses of gauge bosons and those of matter states in a T-dual way. Since the scale of particle masses is below the Planck scale, the mass of these bosons is above the Planck scale; at such a scale, we are not anymore allowed to speak of "gauge bosons" or, in general, fields, in the way we normally intend them.

monopoles would be those of the four-dimensional space coupled to the same photon field A_μ, namely, configurations à la t'Hooft and Polyakov or similar [8]. However, any such topological configuration is characterised by its being living in an infinitely-extended space: only in this way it is in fact possible to make compatible the existence of a p-form working as a "potential" $A_{(p)}$, defined as an analytic function in every point of the space, with the presence of a non-trivial magnetic flux. As is well known, the magnetic flux through a surface can be computed as a loop integral of the vector potential. In the case of a surface enclosing a finite volume, the total flux is the sum of the loop integral circulated in both the opposite directions, so that it always trivially vanishes. However, things are different if the field has a non-trivial behaviour at infinity. At infinity we need just the circulation in one sense, because there is no "outside" from which field lines can "re-enter" in the space: if there is a non-vanishing circulation, there is a non-vanishing magnetic flux, and therefore also a non-vanishing magnetic charge. This however also means that, provided it exists, such a magnetic monopole is a highly non-localised object, with a magnetic field/vector potential such that the magnetic flux vanishes through any compact finite closed surface [9]. As a consequence, also the magnetic charge density vanishes point-wise at any place in the "bulk". Therefore, in our setup, where space is compact, these monopoles cannot exist. Moreover, in our case we don't have a Higgs mechanism either, and, since the surface at infinity does not belong to any configuration of space-time, there is no smooth limit with a true restoration of the conditions at infinity allowing the existence of non-trivial topologies and homotopy groups. Light states with topological magnetic charges do not exist at all, not even approximately as the

[8] for a review and references, see for instance [61, 62].

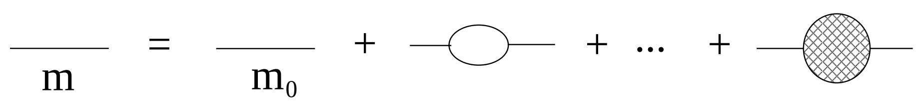
[9] Notice that the situation around the zero-dimensional point is equivalent to the one around the surface at infinity: if on one side the Dirac string can be considered as somehow the "dual" picture of the surface at infinity of the t'Hooft and Polyakov construction, in our scenario both infinity and the dimensionless point are excluded. Differential geometry and gauge theory are here only approximations.

time becomes very large [10].

4.2 Masses and couplings

4.2.1 The mass of a particle

We consider now the masses of the elementary particles. In the string representation, masses arise as ground momenta associated to the states of the string spectrum in a shifted, i.e. contracted, space. Since through 3.1.4 the string scenario is a representation of the combinatorial one, even in the string space a mass is related to the weight of a certain state in the phase space. In the ordinary perturbative approach to field theory (no matter whether it is string-inspired, string-derived, or not) masses, after they have been introduced via some mechanism (Higgs mechanism), are attributes which in general receive corrections at various perturbative orders. The corrections appear as the sum of a series of insertions in the free propagator:



(4.2.1)

Mass and volume in the phase space are related by the fact that the more are the decay channels of a particle, the larger is its entropy, and also the correction to the mass, because higher is the number of virtual processes contributing to the mass renormalization. Heavier particles possess a larger decay phase space: quarks are heavier than leptons, and among leptons neutrinos are the lightest particles. Inside each fa-

[10] The situation is similar to the case of the volume of the group of translations and its identification with the regularized volume of space in the usual normalization of operators and amplitudes, completely absent in our scenario, something that leads to a different interpretation of string amplitudes as global quantities instead of densities, cfr. section 3.1.3.

mily of particles, the heavier (for instance the top as compared to the bottom of an $SU(2)$ doublet) has the larger absolute value of the electroweak charge. In each family, the lightest particle is the one which has less interactions, or less charge (and therefore a lower interaction probability). For instance, $|Q_\nu| < |Q_e|$, $|Q_b = -1/3| < |Q_t = +2/3|$, and quarks, that feel also the $SU(3)$ interactions, are heavier than leptons [11]. Along this line, we can view the lightest particle as the end-point of a chain of projections that reduce the symmetries of the internal space. Heavier particles are therefore those which "occupy" a larger space; they correspond to a larger internal symmetry than lighter particles. Lighter particles correspond to sub-volumes, sub-spaces of those of the heavier particles: the phase space of lighter particles is contained in the phase space of heavier ones. To figure out this point, consider for instance the case of a heavy particle that decays into lighter ones: the physics of these latter is "contained", in the sense that it is produced, derived, by the physics of the heavier one. In terms of combinations of distribution of energies, this simply means that the ways of distributing an amount of energy E along space include as a subset the ways we can distribute an amount $E' < E$. Since mass ratios are related to ratios of occupation volumes in the phase space, and volumes are related to the amount of symmetry, mass ratios turn out to be related to the strength of broken symmetry groups. In our scenario, this is by definition the coupling of the group. For instance, if a mass gap is generated by the breaking of an $SU(2)$ symmetry factor, the mass ratio will be given by the strength of $\alpha_{SU(2)}$. We will derive mass relations in the logarithmic picture (section 4.2.1.5), in which the multiplicative structure of the phase space symmetry groups is mapped into the additive structure of algebras. Instead of couplings one works with the so-called beta-function coefficients. Group ratios result there in differences of beta-function coefficients. In our scenario, the strength $\alpha(G)$ is by definition proportional to the volume of the group, $||G||$ (not to be confused with the volume of the Lie algebra

[11] The first quark family apparently makes an exception: the down quark is heavier, although less charged, than the up quark. This issue will be discussed in detail in section 4.3.2.3.

$||\mathfrak{g}||$), and we can write:

$$\frac{\alpha(G_i)}{\alpha(G_j)} = \frac{||G_i||}{||G_j||}. \tag{4.2.2}$$

Since masses are related to volumes of symmetries, we can write a similar expression:

$$\frac{m_k}{m_\ell} = \frac{||G_k||}{||G_\ell||}. \tag{4.2.3}$$

By comparison of these two expressions we obtain:

$$\frac{m_k}{m_\ell} = \frac{\alpha(G_k)}{\alpha(G_\ell)}. \tag{4.2.4}$$

This expression can also be written as:

$$\frac{m_i}{m_j} = \alpha(G_{ij}) = ||G_{ij}||, \tag{4.2.5}$$

where G_{ij} is a coset. In the logarithmic picture the couplings read:

$$\left.\frac{1}{\alpha_i}\right|_{\log} = \frac{1}{\alpha_0} + \beta_i \ln \mu, \tag{4.2.6}$$

where $\mu = R$ is the scale of space, to be eventually identified with the age of the universe $\mathcal{T}$. Ratios become differences, and we can write:

$$\frac{\alpha_i}{\alpha_j} \to \left.\frac{1}{\alpha_i}\right|_{\log} - \left.\frac{1}{\alpha_j}\right|_{\log} = (\beta_i - \beta_j) \ln \mu, \tag{4.2.7}$$

where β_i, β_j, are the volumes of the symmetry groups G_i, G_j in the logarithmic representation. In a context of group of renormalization, we would call them the beta-function coefficients of the symmetry groups. Since all couplings unify at the Planck scale, in expression 4.2.7 we have considered the additive bare value α_0 to be the same for all of them. This holds if we identify μ with $\mathcal{T}$, the age of the universe. Pulled back to the exponential picture the ratios of masses become then:

$$\frac{m_i}{m_j} = \alpha(G_{ij}) = \mathcal{T}^{\beta_i - \beta_j}. \tag{4.2.8}$$

In order to obtain the masses, we must therefore obtain the "beta-functions" β_i, β_j. In the following, we will proceed to a detailed evaluation of the mass-gap relations as obtained from the stapling of configurations in the logarithmic picture, and relate them to symmetry breaking factors. The Z_2 orbifold pattern through which the staple of configurations is obtained allows to identify as elementary ingredient of all mass relations the coupling of an $SU(2)$ symmetry, not to be confused with the coupling of the $SU(2)_{\mathrm{L}}$ symmetry of the weak interactions: they turn out to be related, but, as the two symmetries are differently defined ($SU(2)_{\mathrm{L}}$ acts chirally on the states), also the coupling strength is different. In first approximation, all the mass gaps can be reduced in different proportions to this elementary step (the approximation is due to the fact that we try to derive masses of free particles, in a scenario in which part of the spectrum is effectively strongly coupled). Once obtained the beta-function coefficient of this elementary block, the coupling strength will be obtained by pulling it back, through exponentiation, to the physical picture.

4.2.1.1 The $SU(2)$ coupling

In order to compute masses, what we need to know is the beta-function of the broken $SU(2)$ group which constitutes the basic ingredient of mass ratios. The beta-function coefficient obtained in the logarithmic picture will become an exponent, i.e. the power to which the radius of the space from the observer up to the horizon (or, equivalently, the age of the universe) must be raised in order to obtain the expression of the effective coupling. In order to determine the $SU(2)$ beta-function, we will derive the volume occupied by the broken $SU(2)$ by counting the volume reductions produced by the various projections we have applied in order to reach the configuration of minimal symmetry. Since our scope is to count a volume fraction within the (massive) matter sector of the physical configuration, the counting must not be done over the full range of string constructions, but just over the span of the massive, physical constructions. Therefore, the effective counting goes from the very latest Z_2 shifts, the one breaking the $SU(2)$ of weak interactions to just one chiral factor, $SU(2)_{\rm L}$, and the one producing masses for all the matter sector, thereby combining left and right moving part of each fermion into one single state, up to the $\mathcal{N}_4 = 2$ point, where the gauge beta-functions vanish. This latter sets the upper bound of the range of projections, because it is starting from this that, by further orbifolding, supersymmetry is broken (entirely, not just partially, broken), leading to a non-vanishing ground energy and therefore a non-flat geometry. At the $\mathcal{N}_4 = 2$ level, which is here the minimal really fully supersymmetric level, even in the presence of shifts along the space-time coordinates masses would be physically irrelevant, because any contribution of a massive particle would be cancelled by the contribution of its superpartners. The projections that effectively produce the elementary spectrum are therefore:

i) the twist that breaks supersymmetry from $\mathcal{N}_4 = 2$ to an apparent $\mathcal{N}_4 = 1$, raising the number of families from one to three, and at the same time breaking the symmetry between leptons and quarks by confining the latter ones, breaking thereby supersymmetry to an effective $\mathcal{N}_4 = 0$. This entails an $SU(2)$ breaking factor, because the **4** containing the lepton as **1** and the quarks as **3** is indeed broken in two parts, as $\mathbf{1} + \mathbf{1_3}$, being the degrees of freedom of the **3** confined into one single strongly coupled state, which, for the electro-weak group, effectively has the same transformation properties of an "up-down" lepton pair;

ii) the four independent rank-reducing shifts that produce the **4** out of the **16** in each family (the third family corresponds to a sector given by the product of projections, therefore it is not the result of independent operations);

iii) the two shifts along the transverse space-time coordinates. Notice that there is no single string orbifold in which all these operations act at the same time: the two shifts on the space-time are applied to different constructions. Nevertheless, the spectrum is produced by their stapling, and therefore is the result of the combined action of the two operations.

This makes in total seven projections. This means that the logarithmic volume of a broken $SU(2)$ factor is $1/7$ of the volume of the initial symmetry of the matter sector. The overall volume of the matter sector is however not the entire volume of space, but just a fraction of it. Space-time is in fact *effectively* doubly-shifted. The Z_2 shift along the space coordinate halves in fact the space (in the logarithmic picture it produces a factor $1/2$, which, when pulled back, i.e. exponentiated, to the physical coordinates produces a square-root contraction of space, $R \to \sqrt{R}$). But the matter sector is produced as the staple of two main orbifold configurations: the series of orbifolds with twisted sectors, that in section 4.1.1.4 we have indicated as "t"-sectors (the entire family of (t,t,t), (t,t,s) and (t,s,s) orbifolds, which descend from the $\mathcal{N}_4 = 2$ orbifold with twisted sector, t) and the orbifold without twisted sectors (the (s,s,s) orbifold, descending from the $\mathcal{N}_4 = 2$, "s" orbifold, in which the twist is associated to a shift along an internal coordinate). The phase space of the matter sector is therefore halved already at the $\mathcal{N}_4 = 2$ level. When stapled, the two configurations give rise to a massive matter sector stapled onto a massive ground (the (s,s,s) orbifold), which gives a $\frac{1}{2} \ln R$ mass contribution (i.e. a $1/\sqrt{R}$ mass factor) common to all matter states. The $SU(2)$ projections differentiating the various matter states split therefore $1/2$ of $1/2$ of the whole space, producing a series of masses that range from just above $1/R^{1/4}$ (that is, $\frac{1}{2}(s) + \frac{1}{2}\frac{1}{7}(t)$) to almost $1/R^{1/2}$ (that is, $\frac{1}{2}(s) + \frac{1}{2}\frac{7}{7}(t)$), i.e., from just above $\frac{1}{2}[\frac{1}{2}(\ln R)]_{(s)}$ to $\frac{1}{2}[\frac{1}{2}(\ln R)]_{(s)} + \frac{1}{2}[\frac{7}{7}\frac{1}{2}(\ln R)]_{(t)}$. The beta-function coefficient of $SU(2)$ is then $\frac{1}{7}$ of $\frac{1}{4}$ of the volume coefficient of the whole space: each $SU(2)$ has therefore a logarithmic volume:

$$\alpha_{SU(2)}|_{\log} \; = \; -\beta_{SU(2)} \ln R \,, \tag{4.2.9}$$

with:

$$\beta_{SU(2)} \; = \; \frac{1}{7} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{28} \,. \tag{4.2.10}$$

The coupling of $SU(2)$ is therefore:

$$\alpha_{SU(2)} \; = \; \mathcal{T}^{-\frac{1}{28}} \,, \tag{4.2.11}$$

where $\mathcal{T} = R$ is the age of the universe. Using the value of the age of the universe given in appendix (eq. A.1), we obtain that, at the present time, $\alpha^{-1}_{SU(2)} \sim 147$. If, to be more precise, we use the age of the universe suggested by the agreement with neutron mass, eq. 4.3.28 (i.e. $\sim 5.038816199 \times 10^{60} \mathrm{M}_{\mathrm{P}}^{-1}$, see appendix), we obtain:

$$\alpha^{-1}_{SU(2)} \sim 147.2 \; (147.211014) \, . \tag{4.2.12}$$

4.2.1.2 The $U(1)_\gamma$ coupling

We have determined the coupling of the symmetry $SU(2)$ from its volume in the phase space, by counting the projections that produce a factorization into $SU(2)$ factors. For this, we did not need to think of $SU(2)$ as of a gauge symmetry. For instance, the factorization of the matter spectrum into three families does not explicitly derive from the breaking of a larger gauge symmetry rotating all the matter states: in the orbifold construction, as soon as extra families show up, they appear as already separated, with their own internal gauge symmetry that replicates the one of the first family. On the other hand, as long as all these states are massless, it is legitimate to think that a higher symmetry should be at work. The orbifold construction represents a phase in which the extra gauge bosons are already projected out from the spectrum. Nonetheless, it turns out useful to think of the $SU(2)$ module of the pattern of symmetry breaking as deriving from the breaking of a larger symmetry, in which all the $SU(2)$ factors are embedded. This allows us to obtain the electromagnetic and weak couplings, explicitly corresponding the one to an unbroken, the other to a broken, gauge symmetry, by comparison with the volume of the $SU(2)$, thought of as deriving from the same overall gauge symmetry. We can obtain the coupling of $U(1)_\gamma$ by determining the ratio of the $U(1)_\gamma$ and $SU(2)$ beta-function coefficients by counting the number of matter states and gauge bosons concerned by the two symmetries. In this way, we don't need to determine the absolute fraction of a group factor within the full symmetry group. The higher is the amount of matter states which are acted on by the symmetry group, the higher is its volume of occupation in the phase space. On the other hand,

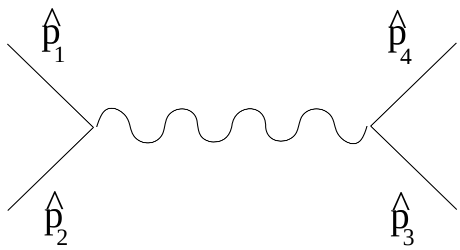


Figure 4.3: Of the four particles involved in the interaction, only three have independent momenta, because the gauge boson "transfers" the condition on the momentum from particles 1 and 2 to particles 3 and 4, imposing $\hat{p}_1 + \hat{p}_2 = \hat{p}_3 + \hat{p}_4$ through the $\delta(\hat{p})$ functions at the interaction's vertices.

gauge bosons act as constraints that reduce the amount of degrees of freedom in the phase space of the matter states: if we have N matter states related by a symmetry carried on by M bosons, $N - M$ matter states have a four-momentum which is not independent, because it is related, through the interaction propagated by a boson, to the one of another matter state. The situation is illustrated in figure 4.3 for the simple case of four particles and one boson, but it can be easily generalized. The beta-function coefficients depend therefore linearly on $N - M$:

$$b_G = \text{const} \times (N - M) \,. \tag{4.2.13}$$

For $N = M$ the beta-function coefficient vanishes. In this framework, since the expressions of the couplings are valid at any scale, therefore up to the Planck scale, this means that the coupling is always 1, at any scale. Indeed, $N = M$ means that there are as many particles as constraints: there are therefore no free degrees of freedom and the phase space volume collapses to one [12]. In the case of $SU(2)$, the

[12] In the logarithmic picture the constructions in which the weakly coupled gauge group appears as perturbative are effectively supersymmetric, with $\mathcal{N}_4 = 2$ extended supersymmetry. The logarithmic picture is in fact obtained through an artificial decompactification of the coupling of the theory. As seen from the logarithmic picture, the beta-function exponent is a $\mathcal{N}_4 = 2$ beta function coefficient. In this case, expression 4.2.13 corresponds to $b = T(R) - C(G)$. An equal number of matter states and gauge bosons, transforming in the same representation, corresponds to an effective $\mathcal{N}_4 = 4$ restoration, a situation of

coefficient has been determined by counting the amount of projections (section 4.2.1.1). This computation already accounts for the factor $(N-M)$, because the projections uniquely determine also the number of states. In order to derive the coefficient of $U(1)_\gamma$, we just need to consider the ratio to the one of $SU(2)$:

$$\frac{b_{U(1)_\gamma}}{b_{SU(2)}} = \frac{(N-M)_{U(1)_\gamma}}{(N-M)_{SU(2)}} \,. \qquad (4.2.14)$$

When counting N and M we must consider all the states as massless (mass gaps are determined as functions of volumes of symmetries. In first approximation we start therefore by considering all the states massless). What enters in the computation of the corresponding beta-function coefficient is therefore the volume of a symmetry as computed by considering all the degrees of freedom as referred to massless states. The beta-function coefficient of $U(1)_\gamma$ is proportional to: 3(families) $\times$ 2($SU(2)$doublets) $\times$ $(\mathbf{1}+\mathbf{3})$(leptons + quarks) $\times$ 2(left + right chirality) $[=48]$ $-$ 1 (gauge boson) $=47$. Notice that, in the counting, we have considered that *all* the matter states are charged under $U(1)_\gamma$. Indeed, three states, the three neutrinos, are uncharged. However, the electromagnetic charge is simply "shifted" from the central value $(\frac{1}{2},-\frac{1}{2})$, but the traceless condition is preserved. As a result, the charge is only "rearranged" among the states: some states result more charged, some less. In total, the strength of the renormalization is the same as with a traceless $U(1)$ with a charge equally distributed among all the states. This is true in first approximation, when all masses are considered vanishing.

The beta-function coefficient of $SU(2)$ is proportional to 48 (the same effective number of states as for $U(1)_\gamma$) minus 3 (the number of gauge bosons), i.e. 45, where the coefficient of proportionality is the

non-renormalization, with vanishing beta-function exponent.

same as for $U(1)_\gamma$ [13]. The ratio of the two coefficients is therefore:

$$\frac{b_{U(1)_\gamma}}{b_{SU(2)}} = \frac{47}{45}\,. \tag{4.2.15}$$

Using 4.2.11 and 4.2.15, and the scale $\mu = \mathcal{T} \sim 5.038816199 \times 10^{60} \mathrm{M}_{\mathrm{P}}^{-1}$, the present age of the universe 4.3.28, adjusted on the neutron mass, we get:

$$\alpha_\gamma^{-1} \sim 183.777867\,. \tag{4.2.16}$$

This has to be considered as a "bare" value of the coupling, not an effective coupling in the field theory sense. We will discuss in section 4.3.3 how this value should be "run back" to obtain the effective coupling to be compared with the value experimentally measured at a certain scale.

4.2.1.3 The $SU(2)_W$ coupling

The strength of the $SU(2)_W = SU(2)_{\mathrm{L}}$ coupling results from the superposition of the various situations in which this interaction appears. As we have seen in section 4.2.1.1, the configuration of minimal symmetry results from the superposition of a geometry in which survives a chiral $SU(2)$ symmetry, that we identify with $SU(2)_{\mathrm{L}}$, and a geometry in which this symmetry is absent and the surviving matter degrees of freedom transform in a $\mathbf{2}$ representation which is transverse to the $\mathbf{2}_{\mathrm{L}}$. The beta-function coefficient of $SU(2)$ was determined by considering the matter space as effectively double-shifted. In the case of the weak

[13] There is here a subtlety: the $SU(2)$ we are here considering corresponds to the smallest $SU(2)$ factor, as resulting from the maximal amount of symmetry breaking. It is therefore basically equivalent to $SU(2)_{\mathrm{L}}$, were not for the fact that the phase space volume of $SU(2)_{\mathrm{L}}$ is enhanced by the fact that it picks a contribution also in the massive, broken-symmetry case, whereas the $SU(2)$ we are here considering refers only to massless states. For what matters $U(1)_\gamma$, it certainly rotates *all* the states, i.e. both left and right movers of any matter states, but, owing to the mass lift of these states, the degrees of freedom of left and right movers are paired, so that they effectively feel the same phase space volume reduction as in the case of $SU(2)$. For practical purposes, it is therefore equivalent to consider the spectrum as massless in both the $SU(2)$ and $U(1)_\gamma$ case, and in the counting of matter degrees of freedom consider also $SU(2)$ as rotating the full spectrum of matter states.

coupling, the beta-function coefficient will result to be a bit higher, because in this case we consider also the phase of massive matter states and gauge bosons. In our scenario, we don't have a Higgs mechanism of spontaneous symmetry breaking. There are therefore no extra states enabling us to formally maintain the same number of states as in a massless situation, and work as everything was massless, by referring the introduction of masses to a second order effect, through the interaction with the Higgs field. The only thing we must do here is to look at the phase space, and compare the situation without and with masses. When the $\mathbf{2}_{\mathrm{L}}$ is broken, we observe an effective halving of the amount of degrees of freedom charged under $SU(2)_{\mathrm{L}}$. Since $SU(2)_{\mathrm{L}}$ rotates doublets, this means that as a symmetry $SU(2)_{\mathrm{L}}$ is broken. This on the other hand is precisely what we expect to observe. A pure counting of degrees of freedom tells us that the beta-function coefficient, given as the average of the two situations, is therefore:

$$\begin{aligned} b_{SU(2)_{\mathrm{L}}} &= \frac{1}{2}\left(b_{SU(2)_{\mathrm{L}}}|_{\mathrm{unbroken}} + b_{SU(2)_{\mathrm{L}}}|_{\mathrm{broken}}\right) \\ &= \frac{1}{2}\left((1)_{\;(\mathrm{unbroken})} + \left(\frac{1}{2}\right)_{\;(\mathrm{broken})}\right) \times b_{SU(2)}\,. \end{aligned} \tag{4.2.17}$$

Inserting the value 4.2.10 of $b_{SU(2)}$ we obtain:

$$\beta_{SU(2)_W} = \frac{1}{2}\left(1 + \frac{1}{2}\right) \times \frac{1}{28}\,. \tag{4.2.18}$$

The present-day value of the inverse of the $SU(2)_W$ coupling is therefore:

$$\alpha_W^{-1} \approx \mathcal{T}_0^{-\left(\beta_{SU(2)_W}\right)} \sim 42.26\,, \tag{4.2.19}$$

where we have used the estimate of the age of the universe given in the appendix, expression A.1. The value 4.2.19 is roughly a factor 4.4 smaller than the inverse electromagnetic coupling, given in 4.2.16. Also this number has to be considered a "bare" value, to be corrected in the way we will discuss in section 4.3.3.

4.2.1.4 The strong coupling

Expression 4.2.13 holds only as long as the gauge interaction is weakly coupled. This means that we cannot use the difference between number of matter states and number of gauge bosons as a discriminant in order to say whether a given gauge group is confining or not. As we have seen, $N - M$ corresponds to an ordinary concept of gauge beta function coefficient only in an effective $\mathcal{N}_4 = 2$ representation of physics, where this reflects the $b = T(R) - C(G)$ expression of the beta-function coefficient. If blindly applied, this computation would imply that $SU(3)_c$ is not strongly coupled. Nevertheless, an investigation of the $\mathcal{N}_4 = 0$ gauge beta function tells us that $SU(3)_c$ is confining. But what does it really mean "confining"? Experimental investigations say that at the scale of some typical quark process, for instance the Z-boson mass in a $e^+e^- \to 4\mathrm{J}$ event, $\alpha_s(M_Z) = 0.119$ [63]. This means that the coupling is definitely stronger than the electromagnetic coupling, therefore justifying the fact that the proton "holds up" despite the repulsive electromagnetic force acting among its quarks, yet it has anyway a value lower than one, therefore not really what in our theoretical framework we call "strong coupling". Indeed, it is weak enough to allow obtaining a glimpse into the parton structure, because the quarks are not so tightly bound to appear like just one single, elementary state. In our derivation of the set of minimal symmetry configurations, we have seen that the quark sector feels an internal force which is at the strong coupling, i.e. it has a coupling strength larger than 1. As seen from a picture in which $U(1)_\gamma$ and $SU(2)_W$ are at the weak coupling, the internal symmetry appears as a symmetry relating three families of three quark colours. Therefore, strictly speaking an $SU(9)$ symmetry, which leads to the existence not only of colour singlets formed with quarks belonging to one single family, but also with quarks belonging to different families. Of this type are for instance the D-mesons, formed by coupling charm and down quarks. However, the splitting into three matter sectors does not necessarily imply that any single sector is a replica with its own gauge bosons: any time we invert the game by going to an S-dual picture, where we switch on the gauge part as a weakly coupled gauge group of which we explicitly see the bosons as massless string states (e.g. the heterotic picture) we see only three matter sectors all charged under one single gauge group. Therefore, either i) we see the sector as strongly coupled, a situation in which it does not make any sense to speak of gauge bosons, because there are no gauge bosons at all being the group at confinement, or ii) we have a perturbative realization in which we explicitly see gauge bosons, but in this case there is one single gauge group sector, whose index is beared by all the matter states. In this second case, could we see the internal group as explicitly realized, we would see an $SU(3)$ index beared by the three colour states of each family. The real physical situation is however the one of a basic strong coupling, in which, owing to compactness of space-time, T-duality, although "entropically" broken, plays a fundamental role. The ground strength of the coupling of the theory is set by the gravitational coupling, which in our theoretical framework is set to 1, as the unit in which everything else is measured. Its decoupling is only an artifact introduced in order to investigate the theory content in a flat space (logarithmic picture), where gravity, and the local geometry of space-time, is factored-out, and we work in a flat space. Indeed, all the other couplings depend on the coordinates of space-time (in practice, on the age of the universe), and strong and weak coupling correspond to situations which are T-dual, where the turning point is precisely the gravitational coupling. It is precisely the non-complete disappearance of T-duality what allows to speak in terms of minimal length also in the description on the continuum, even after symmetry has been broken by the stapling of all the possible geometries. This on the other hand implies that physics always results from the stapling of T-dual (or, to better say, S-dual) strong and weak coupling phases. In the case of the colour interaction, this means that, under certain conditions, we can detect certain properties that we can interpret as belonging to a weak coupling phase (i.e. $\alpha_{SU(3)} < 1$). The strength of the coupling for the weak coupling phase is derived as in section 4.2.1.2, evaluating the volume in the phase space by counting the number of matter states and gauge bosons, $N - M$. We have

$N = 3$ (quark colours) $\times$ 3 (families) $\times$ 2 ($SU(2)_W$ indices) $\times$ 2 (left + right chirality) $=$ 36, and $M = 3^2 - 1 = 8$ gauge bosons. Therefore, $N - M = 28$. The strength of the $SU(3)$ coupling is therefore computed form that of $SU(2)$ as in 4.2.14:

$$\frac{b_{SU(3)_c}}{b_{SU(2)}} = \frac{28}{45} \,. \tag{4.2.20}$$

This implies:

$$\alpha^{\mathrm{T}}_{SU(3)} = \mathcal{T}^{-\frac{1}{45}} \,, \tag{4.2.21}$$

where we have indicated with "T" the fact that we are considering the T-, or more precisely the S-, dual of the coupling. Inserting the value of the age of the universe, expression A.1, we obtain:

$$\alpha^{\mathrm{T}}_{SU(3)} \sim 0.0448 \,. \tag{4.2.22}$$

4.2.1.5 Elementary masses

We will proceed now to a determination of the mass ratios, as functions of ratios of symmetry volumes. Since these relations hold at any scale, we may think of working at a time scale close to the Planck scale, and map to a logarithmic picture, in which all the couplings are very small (remember that they go to 1 to ward the Planck scale. Their logarithm therefore vanishes). This procedure will give us a first estimate of the mass relations, as functions of just one coupling, the $SU(2)$ coupling. These relations produce reasonable mass values for the stable particles, whereas for the unstable, and heavier, ones, the phase space volumes are more strongly affected by the superposition with other energy scales, and, in order to produce values comparable with experiments, a more detailed knowledge of the dynamics and the experimental conditions is required. The actual computation of current mass values will be considered in section 4.3.

According to what discussed in section 4.1.1.4, a pure $SU(2)$ symmetry factor is the distance separating the first from the second neutrino: it is in fact a simple passage from a certain amount of degrees of freedom to exactly twice as much, while keeping fixed the type of interaction the particles are sensitive to. Moreover, it involves the most "neutral", i.e. less interacting, particles of the spectrum. In particular, these particles do not feel the strong interaction. As the volume of a particle in the phase space is related to the amount of interactions the particle is involved in, we expect the first two neutrinos to be also the least affected ones by perturbations and corrections due to an imprecise evaluation of the whole dynamics. Working in the logarithmic picture, more than in ratios we are interested in differences. Close to the Planck scale, all interactions are strong, and the spectrum tends to arrange into compounds neutral for the various interactions. Since we investigate the states in a logarithmic representation, we treat them nevertheless as free states. However, of interest for us is the hierarchy underlying the formation of neutral compounds. As the electromagnetic interaction is stronger than the weak interaction, the spectrum organizes in a way to first separate into electrically neutral states, and then it arranges into $SU(2)_W$ doublets. This means in particular that all the neutrinos are lighter than any other particle, and constitute the first three lightest steps in the mass hierarchy. From expression 4.1.8 we see that the distance between the first neutrino (the lightest, ν_e) and the ground momentum is $\frac{5}{3}$, the distance from the second (ν_μ) to the first neutrino is $\frac{3}{2}$, and the distance from the third (ν_τ) to the second is 1. Let us call these distances $\delta^{(1)}$, $\delta^{(2)}$ and $\delta^{(3)}$ respectively, and α the strength of the $SU(2)$ coupling. $\delta^{(2)}$ corresponds then to α^{-1}:

$$\begin{aligned} \delta^{(1)} &= \alpha^{-1 \times \frac{5}{3}\frac{2}{3}}\,; \\ \delta^{(2)} &= \alpha^{-1}\,; \\ \delta^{(3)} &= \alpha^{-1 \times \frac{2}{3}}\,; \end{aligned} \tag{4.2.23}$$

In order to simplify the following discussion, let us introduce the notation $\delta^{(2)} = \ln \alpha^{-1} \equiv \ln \delta$. We can therefore write the following relations, holding in the logarithmic picture:

$$\begin{aligned} m_{\nu_e} - \ln M_0 \equiv \delta^{(1)} &= \frac{10}{9} \ln \delta \, ; \\ m_{\nu_\mu} - m_{\nu_e} \equiv \delta^{(2)} &= \ln \delta \, ; \\ m_{\nu_\tau} - m_{\nu_\mu} \equiv \delta^{(3)} &= \frac{2}{3} \ln \delta \, , \end{aligned} \tag{4.2.24}$$

where $M_0 = 1/2\sqrt{\mathcal{T}}$. Physical masses are related therefore through the following ratios:

$$\frac{m_{\nu_e}}{M_0} = \delta^{\frac{10}{9}} \, , \tag{4.2.25}$$

$$\frac{m_{\nu_\mu}}{m_{\nu_e}} = \delta \, , \tag{4.2.26}$$

$$\frac{m_{\nu_\tau}}{m_{\nu_\mu}} = \delta^{\frac{2}{3}} \, . \tag{4.2.27}$$

Let us now consider the electrically charged matter states. Each family contains a lepton and quarks, with electrical charges that make any family overall electrically neutral in itself. That is, the charged counterpart of each family is overall neutral. Differently to the field theoretical approach, our scenario, being defined for any value of the age of the universe, is valid at any energy scale till the Planck scale. We can therefore consider the situation toward the Planck scale, where also the electromagnetic interaction is strong. In this regime, the charged states of each family glue together to form an electrically neutral compound. From the point of view of the phase space, this compound behaves therefore similarly to the corresponding neutrino.

Indeed, behaving effectively as a single particle, we expect that the sets of the charged particles of each family stay in the same relative ratios (or logarithmic distance, if one prefers) as the inter-family ratios (distances) of neutrinos. Do they have also the same mass? Not at all, because of the higher amount of "internal" degrees of freedom, that come into play as soon as, lowering the energy scale, they get "unfrozen", freed up into independent degrees of freedom, that occupy therefore a larger volume in the phase space. If we indicate with $\mathcal{V}(q_{\rm up}, q_{\rm down}, \ell)$ the volume of the charged part of each family, we expect therefore:

$$\mathcal{V}(u,d,e) \;\sim\; \Delta^{\frac{10}{9}}\,; \tag{4.2.28}$$

$$\mathcal{V}(c,s,\mu) \;\sim\; \Delta^{1+\frac{10}{9}}\,; \tag{4.2.29}$$

$$\mathcal{V}(t,b,\tau) \;\sim\; \Delta^{\frac{2}{3}+1+\frac{10}{9}}\,. \tag{4.2.30}$$

In this case, the ground mass factor is the volume of the neutrino sector, i.e. the mass of the heaviest neutrino, m_{ν_τ}. The fact that toward the Planck scale these states are effectively at the strong electromagnetic coupling, and therefore exist only as singlets, implies in the logarithmic picture a reduction of their symmetry span by a factor $\frac{1}{3}$. Once pulled back to the physical picture, this results in a third-root power contraction of the phase space volume of the charged part of each family.

Furthermore, differently from what happens for the leptons, in the case of quarks the $\alpha_{SU(2)}$ factor between the top and bottom quark does not separate the masses of the single quarks, but singlets of the confining $SU(3)$ symmetry. Depending on whether the latter are formed by pairing two quarks (quark-antiquark pair, like in the mesons) or three quarks, like in the proton or the neutron, we expect therefore a normalization factor of about 1/2 or 1/3. Differently from what happens for the other symmetry groups, in the logarithmic picture the confining symmetry is not perturbatively realized. We expect therefore the normalization coefficients to not appear at the logarithmic level, to be promoted to exponents of the age of the universe, but to be true multiplicative normalization factors.

In the case of the quarks of the first family, the so called experimental values are quantities derived rather indirectly. Contrary to the second and third family, where the mass values are obtained from the mass of the unstable particles they form basically by pairing two by two (mesons), in this case we only detect particles (pions, and hadrons) strongly affected by the neutron mass scale. We give therefore here the expression for the "bare" quark mass, without introducing normalization factors, because the contact with experiments does not occur at the bare quark level. We expect the up-to-down mass relation to be:

$$\max\left[m_u, m_d\right] \;\sim\; \delta^{\frac{1}{3}} \times \min\left[m_u, m_d\right] . \tag{4.2.31}$$

Here we have indicated in brackets the heavier and the lighter mass of the up and down quark pair. In principle the first should be the up quark mass, and the second the down quark mass, however, as we will explain in section 4.3.2.3, for the first quark generation they turn out to be exchanged.

The mass gap between quarks and leptons is the consequence of the breaking of the **4** of each family into $\mathbf{1} \oplus \mathbf{3}$. This separates the phase space into two parts of unequal volumes. Counting the weights in the logarithmic picture, we see that the weight of the **1** is one-half of the **2** occurring when the **4** of $SU(4)$ is broken in the $\mathbf{2}+\mathbf{2}$ of $SU(2) \times SU(2)$. The distance between the two parts is one-half of the logarithmic volume of $SU(2)$. Therefore, we expect the "up" of the **1** to lie a $\sqrt{\delta} = \sqrt{\alpha_{SU(2)}}$ factor below the "down" of the **3**. Taking then the third root of the broken $SU(2)$ volume fraction, as required by fixing the normalization of volumes at the strong electroweak coupling (at the Planck scale), we obtain the following separation factor between the lighter quark and the electron:

$$\min\left[m_u, m_d\right] \;\sim\; \sqrt{\delta}^{\frac{1}{3}}\, m_e\,, \tag{4.2.32}$$

where, by analogy with 4.2.25,

$$m_e \;\sim\; \delta^{\frac{10}{9}}\, m_{\nu_\tau} . \tag{4.2.33}$$

Putting 4.2.31, 4.2.32 and 4.2.33 together, we obtain:

$$\tilde{\mathcal{V}}(u,d,e) \;\equiv\; \mathcal{V}(u,d,e) \quad\sim\quad \delta^{\frac{1}{3}}\, \sqrt{\delta}^{\frac{1}{3}}\, \delta^{\frac{10}{9}} \quad\sim\quad \Delta^{\frac{10}{9}} . \tag{4.2.34}$$

From these relations we derive then the corresponding factors for the second and third family:

$$m_\mu \sim \delta\, m_e\,, \tag{4.2.35}$$

$$m_\tau \sim \delta^{\frac{2}{3}}\, m_\mu\,. \tag{4.2.36}$$

Similarly, we have:

$$m_s \sim N_s \left(\sqrt{\delta}^{\frac{1}{3}}\right)^{\frac{9}{10}\left(1+\frac{10}{9}\right)} m_\mu\,, \tag{4.2.37}$$

$$m_c \sim N_c \left(\delta^{\frac{1}{3}}\right)^{\frac{9}{10}\left(1+\frac{10}{9}\right)} m_s\,, \tag{4.2.38}$$

$$m_b \sim N_b \,(\sqrt{\delta}^{\frac{1}{3}})^{\frac{9}{10}\times\frac{2}{3}} m_\tau\,, \tag{4.2.39}$$

$$m_t \sim N_t \,(\delta^{\frac{1}{3}})^{\frac{9}{10}\times\frac{2}{3}} m_b\,. \tag{4.2.40}$$

Here we introduced the normalization coefficients N_s, N_c, N_b and N_t in order to account for the fact that the values we want compute refer to colour singlets. For the charm, bottom and top quarks the mostly observed states are unstable particles formed by a quark-antiquark pair, therefore the coefficient N_i, $i = c, b, t$ is expected to be $\frac{1}{2}$. In the case of the s quark, it seems that, besides the K mesons, also the Λ and Σ baryons play a relevant role, so that, as a matter of fact, as we will see in section 4.3.2, things work if we set $N_s = \frac{1}{3}$. A detailed prediction of these coefficients would require a complete analysis of the interactions of the corresponding quark, in order to see, in the light of this theoretical framework, what are the relative weights of the various contributions to the staple of configurations that build up the full phase space of each particle.

The value of the $SU(2)$ coupling, δ, must be always intended as run to the appropriate energy scale. This means that it is not constant through all the mass hierarchy. The reason of this behaviour, which by the way is common also to the weak and electromagnetic coupling, is the following. We measure group volumes within the mass spectrum. However, as we just discussed, the overall mass is not just given by the position of a particle in the hierarchy of broken symmetries within the matter sector, but receives a contribution from the ground momentum. In logarithmic terms, whenever a mass distance of two particles, 1 and 2, related by an $SU(2)$ relation, can be written as $\Delta m = \ln \delta$, the actual value of δ is computed by considering the volume of symmetry breaking as measured in terms of the overall volume. Namely:

$$m_2 = \ln(\delta/V) + m_1^{(0)} + \ln V \,, \tag{4.2.41}$$

where V is the overall ground volume, that includes the ground momentum and the volume of all the particles lighter than particle 1. Indeed, we have here indicated with $m_1^{(0)}$ the mass distance of the lighter particle to the next one in the step-down mass hierarchy. When we consider the whole value of a mass, we must "normalize" the value of the coupling. The higher is the volume V, i.e. the heavier is the lighter mass of the pair, the lower is the distance (remember that $\delta > 1$, being the inverse of a weak coupling). Since we are going to construct the tower of masses step by step starting from the lightest one, by considering distances investigated in the logarithmic picture, what we are building is a series of higher levels in the mass hierarchy in which at any step the coupling is corrected logarithmically. The number given in 4.2.12 is therefore the value of the inverse of the coupling at the M_0 scale introduced on page 142. The value of δ to be inserted in the second-to-first neutrino mass ratio is not the inverse of the "natural" value of the $SU(2)$ coupling, but the value linearly run from M_0 to the neutrino scale, and so on.

4.3 Present-time values of masses and couplings

4.3.1 Bare masses

Now that we have at hand the value of the $SU(2)$ coupling we can proceed to an explicit evaluation of the masses of the elementary particles, as they can be computed using the mass formulae given in section 4.2.1.5. These can be considered the "bare" values. A comparison with the experimental results must take into account the conditions under which a certain mass is operatively defined. In particular, unstable particles (in practice all apart from the particles of the first family, of which however the only relevant case is the electron, because the up and down quark masses are only indirectly measurable) turn out to be strongly affected by the superposition with the stable scale, $m_{3/10}$, that corresponds to the proton/neutron mass scale (see section 4.3.6). We will proceed to the evaluation according to section 4.2.1.5, by inserting the value of δ, the inverse of the $SU(2)$ coupling, recalculated at the appropriate scale through a linear running in the logarithmic picture (a linear running of the logarithm of the coupling), obtained by imposing that at the Planck scale the coupling is zero, and at the mass scale $M_0 = \frac{1}{2}\mathcal{T}^{-\frac{1}{2}}$ it corresponds to the value 4.2.12. We will assume that the coupling renormalizes in correspondence of each $SU(2)$ step. Therefore, we will use the value of δ at the scale M_0 in order to compute the mass of the first neutrino, then we will use the value of δ recalculated at the scale of the first neutrino in order to obtain the mass of the second. For the third we will not recalculate the coupling, because the third generation is not produced by an independent $SU(2)$-breaking projection, but is generated by those that give origin to the first and the second matter generation. Similar considerations apply also to the hierarchy of the charged particles.

4.3.2 Mass values

4.3.2.1 Neutrinos

The first masses we calculate with this method are those of the three neutrinos. Using the value of the present-day age of the universe derived from the neutron mass, expression 4.3.28, from 4.2.24 and 4.2.25 we obtain:

$$M_0 \;=\; 2.23 \times 10^{-31}\,\mathrm{M_P} \;=\; 2.72 \times 10^{-12}\,\mathrm{GeV}\,. \tag{4.3.1}$$

From this, according to 4.2.24 and 4.2.25–4.2.27, we compute:

$$\begin{aligned} m_{\nu_e} &= 1.40\,\mathrm{eV}\,, \\ m_{\nu_\mu} &= 205.25\,\mathrm{eV}\,, \\ m_{\nu_\tau} &= 5.72\,\mathrm{KeV}\,. \end{aligned} \tag{4.3.2}$$

These values agree with the experimental indications of possible neutrino oscillation effects at the electronvolt scale.

4.3.2.2 Electron

The mass of the electron is then derived from 4.2.33 by using the value of δ renormalized at the m_{ν_μ} mass scale. We obtain:

$$m_e \;=\; 0.506\,\mathrm{MeV}\,. \tag{4.3.3}$$

In order to compare this value with the experimental one, we must reproduce the conditions under which this quantity is measured. The electron mass is derived from the Rydberg constant, entering the expression of the energy levels of the atomic emission spectra. The electron which is measured is not therefore a truly free electron but an orbiting electron, which interacts with the proton in an electron+proton +neutron system. Since the interaction with the proton is of electromagnetic type, its strength is set by the electromagnetic coupling α_γ. We can expect that the volume occupied in the phase space is set by the fraction of the proton volume involved in the interaction with the electron, multiplied by the fraction of volume occupied by the electron as compared to the proton. In practice, proportional to the coupling times the volume of the electron times the volume of the proton, measured in units of the volume of the proton:

$$\Delta E \sim \alpha_\gamma \times m_{\rm p} \times \frac{m_e}{m_{\rm p}} \sim \alpha_\gamma \, m_e \,. \tag{4.3.4}$$

This can be viewed as a "quantum gravity" correction to the electron mass. Being measured through atomic spectra means in fact in particular that the electron lives in a space curved by the overall energy of the atomic system. The correction to the ground energy of the electron is expected to be given by the gradient of energy:

$$\Delta E \approx |\nabla E| \,, \tag{4.3.5}$$

In a logarithmic picture, this becomes:

$$E \longrightarrow E|_{0=m} + |\nabla \ln E| \,, \tag{4.3.6}$$

where it is intended that dimensions are adjusted by appropriate powers of c and $\hbar$. This is also the kind of correction considered in ref. [14] (see chapter 7), that modifies the effective value of $\hbar$. The correction is higher the higher is the curvature of space, and vanishes in a flat space. Although strange it may look, this correction term is of the form:

$$\Delta E \sim \frac{1}{(c)\Delta t} \sim \frac{1}{\Delta x} \times \frac{E}{E} \,. \tag{4.3.7}$$

It has therefore the appropriate form in order to represent a quantum correction to the energy. Now, what is the gradient of energy in an atom, in particular a hydrogen atom? The total energy is basically the proton plus neutron mass, which is concentrated in a space region of Bohr radius. Therefore, in units for which $\hbar = c = 1$:

$$\nabla \ln E = \partial_r \ln E \sim \frac{1}{E} \times \frac{E}{r_{\rm Bohr}} \sim \frac{1}{m_{\rm p} + m_{\rm n}} \times (m_{\rm p} + m_{\rm nn}) \times \alpha_\gamma \, m_e \,. \tag{4.3.8}$$

This implies that the electron mass is corrected to:

$$m_e \;=\; m_e + \alpha_\gamma m_e \,, \tag{4.3.9}$$

where α_γ must be run to the center-of-mass scale of the hydrogen atom (see section 4.3.3 for a discussion of the running). We obtain therefore expression 4.3.4. Taking this correction into account, we obtain:

$$m_e \;\sim\; 0.5107\,\mathrm{MeV}\,. \tag{4.3.10}$$

The electron interacts then with the proton and the neutron also at higher orders, through the $SU(2)_W$ weak coupling. These are however very minor corrections that do not change the value of the mass at degree of approximation of expression 4.3.10. The official value reported in the literature is:

$$m_e|_{\mathrm{experimental}} \;\sim\; 0.511 + \mathcal{O}(10^{-4}) \;\;\mathrm{MeV}\,. \tag{4.3.11}$$

At this stage, it does not make sense to look for a better matching of our predictions with this value, because experimental measurements are performed in an indirect way, and depend on the theoretical framework within which the fine corrections are computed. A true comparison with experiments going beyond a first approximation would require implementing and interpreting the whole measurement process within our theoretical scheme.

4.3.2.3 Up and down quarks

Continuing along the lines of section 4.2.1.5, from 4.2.31 and 4.2.32 we obtain:

$$\min[m_u, m_d] \;=\; 1.108\,\mathrm{MeV}\,; \tag{4.3.12}$$
$$\max[m_u, m_d] \;=\; 5.305\,\mathrm{MeV}\,. \tag{4.3.13}$$

The reason why we did not yet specify which one of the quarks we are considering is due to the fact that, although in principle we should find as lighter particle the quark with the lowest electric charge, namely the down quark, the role of up and down quark are exchanged. The explanation has to do with the way in our framework the symmetry breaking is realized. At low energy, the $SU(2)_W$ symmetry appears as a broken gauge symmetry, with the breaking tuned by a parameter of the order of a negative power of the age of the universe. As we will see in section 4.3.5, the $SU(2)_W$ gauge boson masses scale in such a way that $\mathcal{T} \to \infty$ is a limit of approximate restoration of the $SU(2)_W$ symmetry. Moreover, remember that the weak force in itself is stronger than the electromagnetic force: $\alpha_W > \alpha_\gamma$ (it is called weak because for low transferred momenta, $p/M_W \ll 1$, effective scattering/decay amplitudes are suppressed by the boson mass: $\alpha_W^{\rm eff} \approx \alpha_W / M_W$). Therefore the "hierarchy" of matter is prioritarily determined by the $SU(2)_W$ charge, more than by the electric charge. As a consequence, the matter spectrum can be thought of as being made of two subspaces, the "up" and the "down" subspace, and the trace of the electric charge can be viewed as:

$$< Q_{e.m.} > = \sum_{\ell, {\rm q}} < {\rm up} | Q_{e.m.} | {\rm up} > \, + \sum_{\ell, {\rm q}} < {\rm down} | Q_{e.m.} | {\rm down} > , \tag{4.3.14}$$

where $\sum_{\ell, {\rm q}}$ indicates the sum over leptons and quarks. The condition of approximate restoration of the $SU(2)_W$ symmetry, and the dominance of the weak force with respect to the electromagnetic one, require that the two terms of the r.h.s. of 4.3.14 give an equal contribution to the total mean value of the electric charge. Otherwise, this would explicitly break the $SU(2)_W$ invariance. This imposes that the trace of the electric charge has to vanish separately on the "up" and "down" multiplets. In practice, both of them must vanish. For the validity of this argument it is essential that the weak force ends up by dominating the more and more over the electric one, and that the symmetry is restored at infinitely extended space-time; therefore, the full space must be essentially thought of being as separated in two $SU(2)_W$ eigenspaces. Compatibility of the theory at any finite time with the situation at the limit tells us that:

$$\mathrm{tr}\,(\nu, d) \, = \, 0 \, . \tag{4.3.15}$$

Since the ν charge vanishes, we have that:

$$\mathrm{tr}\,(d) \;=\; 0\,. \tag{4.3.16}$$

This is only possible if, for one family, the roles of the up and down quarks, for what matters the electric charge, are exchanged, so that we have $\mathrm{tr}\,(d) \;=\; 3 \,\times\, \left(\frac{2}{3} - \frac{1}{3} - \frac{1}{3}\right) \;=\; 0$. Correspondingly, the trace of the "ups" is also vanishing:

$$\mathrm{tr}\,(\mathrm{e}, \mu, \tau, \mathrm{u}) \;=\; -1 \;-1\; -1 \;+\; 3 \times \left(-\frac{1}{3} + \frac{2}{3} + \frac{2}{3}\right) \;=\; 0\,. \tag{4.3.17}$$

Therefore, in one of the three quark families the role of up and down is interchanged: the quark with electric charge +2/3 is indeed the "down", while the one with charge -1/3 is the "up". In the ordinary field theory approach, this argument does not apply because the symmetry remains broken also at infinitely extended space-time [14]. Simple entropy considerations allow us to identify in which family the flip occurs. Let's consider the $SU(3)_c$-singlet made out of the three quarks, one per each family, with higher electric charge, and the one made in a similar way out of the three quarks with the lower electric charge. Clearly, the first one is the most interacting singlet we can form by picking one quark from each family, and conversely the other one is the less interacting one we can form. The first must therefore also be the most massive out of all the possible $SU(3)$-singlets formed by one quark per each family, while the second one must be the lightest. The only possibility we have to achieve this condition is when the flip between charge +2/3 and -1/3 quarks occurs in the lightest family, i.e., for the quarks we usually call the up quark and the down quark. Therefore, approximately the value of the mass of the up quark is the one we computed for the lightest "down" quark state, and conversely the mass of the down quark is the one we assigned to the lightest "up":

$$m_u \;=\; 1.108\,\mathrm{MeV}\,; \tag{4.3.18}$$
$$m_d \;=\; 5.305\,\mathrm{MeV}\,. \tag{4.3.19}$$

[14] Notice that the usual charge assignment breaks the $SU(2)$ symmetry explicitly.

Possible further minor corrections, that we are not able to estimate here, are to be expected as a consequence of the fact that now the lighter quark has a higher absolute value of the electric charge, and therefore a larger volume in the phase space, whereas the upper quark has a lower absolute value. The up quark could possibly be slightly heavier, and the down quark slightly lighter.

4.3.2.4 The charged particles of the second and third family

Passing to the second family, we first compute the muon mass. From 4.2.35, with a value of δ recalculated at the electron mass scale, we obtain:

$$m_\mu = 55.6\,\mathrm{MeV}\,. \tag{4.3.20}$$

Analogously, by recalculating δ at the m_μ scale, we proceed to compute the strange quark mass from 4.2.37, obtaining:

$$m_s = 94.4\,\mathrm{MeV}\,. \tag{4.3.21}$$

Recalculating δ at the strange quark mass scale, from 4.2.38 we obtain then:

$$m_c = 1.22\,\mathrm{GeV}\,. \tag{4.3.22}$$

The experimental values are: $m_\mu = 105.658\,\mathrm{MeV}$, $m_s \sim 95 \pm 5\,\mathrm{MeV}$ and $m_c \sim 1.29^{+0.05}_{-0.11}\,\mathrm{GeV}$. From 4.2.36, still using the value of δ evaluated at the electron mass scale, we calculate then:

$$m_\tau = 1.27\,\mathrm{GeV}\,, \tag{4.3.23}$$

and, by recalculating the value of δ at the m_c mass scale,

$$m_b = 5.32\,\mathrm{GeV}\,, \tag{4.3.24}$$

and

$$m_t = 186\,\mathrm{GeV}\,. \tag{4.3.25}$$

4.3.2.5 The neutron mass

Now that we know what the spectrum of elementary matter degrees of freedom is, we are in a position to reconsider the neutral average scale introduced in section 3.4. Toward the Planck scale, all interactions tend to the strong coupling with the same strength. The only possible state in the spectrum is therefore a compound neutral for all the three forces. For the first generation, this can only be a compound of neutrino, electron, proton and neutron, and their charge conjugates. This pattern is replicated in all the three families. However, as soon as we depart from the Planck scale toward lower scales, heavier similar compounds tend to decay to lighter states. The only really stable state is therefore the one of the first family. Stability is here to be intended in a temporal sense, i.e. when a large time interval is considered. Otherwise, at the typical electro-weak scales, the components of this state which are neutral for the colour force exist also as free states. On the other hand, since any experimental measurement is performed during a finite time interval, the physical measurements are an average both over the staple of configurations at any time, and over the time duration of the experiment. As a consequence, the shorter is the mean life of unstable particles, the more are their properties "blurred" by the presence of this state.

Since electron and neutrino are very light as compared to its mass scale, we must think that the dominant contribution is given by the colour singlets, the proton and the neutron. The difference between the two consists in an $SU(2)_W$ rotation, that, owing to the much weaker strength of the coupling as compared to α_s, is expected to not affect that much the mass scale of the colour singlet. On the other hand, we may also think that, as we approach the Planck scale, the proton first tends to glue with the electron and neutrino, to form an electromagnetic-neutral compound, and then this in turn glues to the neutron to form a state neutral also under the $SU(2)_{\mathrm{L}}$ interaction. In first approximation, we can therefore assume that the mass scale of the electromagnetic neutral compound (proton, electron, neutrino) is close to the one of the neutron, and the mass obtained in Chapter 3 (eq. 3.4.2) gives four times the mass of the neutron. Therefore:

$$m_{\mathrm{n}} \simeq \frac{1}{4} < m > \simeq \frac{1}{8} \mathcal{T}^{-\frac{3}{10}} . \tag{4.3.26}$$

By inserting in 4.3.26 the current value for the age of the universe, as obtained by extrapolating data of experimental observations within the theoretical framework of Big Bang cosmology, we obtain a value quite close to the neutron mass. Namely, from 4.3.26 and a central value of the age of the universe $\sim 12.75 \times 10^9$ yrs, ($\sim 5 \times 10^{60} \mathrm{M}_{\mathrm{P}}^{-1}$, see appendix) we obtain:

$$m_{\mathrm{n}} \approx 937 \, \mathrm{MeV} , \tag{4.3.27}$$

quite in good agreement with the value experimentally measured, $939.56563 \pm 0.00028 \, \mathrm{MeV}$ [64]. A more correct analysis would require a new derivation of the value of the age of the universe completely *within our framework*. On the other hand, within our theoretical scheme one can reverse the argument, and take the mass of the neutron as the most precise measurement of the age of the universe. In this case, we obtain as its actual value:

$$\mathcal{T}_0 = 12.62028271 \times 10^9 \, \mathrm{yr} . \tag{4.3.28}$$

The fact that our mass formula gives as average mass the mass of the neutron is nicely in agreement with what we would expect from a universe behaving as a black hole. According to the common astrophysical models, a black hole is in fact the step just following the "neutron star" phase of a star at the end of its life. Our considerations of above suggest that the universe can be roughly thought of as a kind of neutron star at the point of transition to a black hole.

4.3.2.6 The proton mass

Proton and neutron differ by an $SU(2)_W$ rotation of the quarks. We expect therefore that the main contribution to the mass difference between the two is set by the mass difference of the up and down quarks. However, since in this case quarks are strongly coupled and confined, their phase space volume is reduced as compared to that of the free quarks, corresponding to the mass values 4.3.18, 4.3.19. From 4.3.18 and 4.3.19 we obtain a mass difference given by:

$$\Delta m_{u-d} \; = \; 4.197\,\mathrm{MeV}\,. \tag{4.3.29}$$

If we think of contracting the phase space volume by a factor $\frac{1}{3}$, as to be expected when we glue three quark degrees of freedom into a singlet, we should expect an effective neutron-proton mass difference given by one third of 4.3.29, namely:

$$m_{\mathrm{n}} - m_{\mathrm{p}} \; \sim \; 1.399\,\mathrm{MeV}\,. \tag{4.3.30}$$

The mass difference experimentally observed is about:

$$(m_{\mathrm{n}} - m_{\mathrm{p}})|_{(\mathrm{experimental})} \; = \; 1.293\,\mathrm{MeV}\,. \tag{4.3.31}$$

If we run this value from the free quarks mass scale to the proton mass scale, by assuming a linear running of masses in the logarithmic picture as for the effective couplings we obtain:

$$(m_{\mathrm{n}} - m_{\mathrm{p}})_{\mathrm{corr.}} \; \sim \; 1.291\,\mathrm{MeV}\,. \tag{4.3.32}$$

We do not insist here on the exact computation of this value, for which a better knowledge of the physical details would be necessary. However, what we learn from this discussion is that the values we find are in principle compatible with the experimental observations.

4.3.3 The effective electromagnetic coupling

The couplings α_γ, α_W and α_s derived in section 4.2.1.1 and 4.2.1.2 and 4.2.1.4 run with time, and therefore with an energy scale: they are the couplings at a specific age of the universe. The values we obtained do not however correspond to the actual value of the physical coupling. In order to obtain the latter, we must run them up to the appropriate scale. In this section we consider the correction to the weak couplings.

In the representation in which elementary particles are defined, namely in the logarithmic picture, the effective gauge couplings are corrected according to:

$$\alpha_i^{-1} \approx \alpha_i^{-1}|_0 + b_i \ln \mu / \mu_0 \,, \tag{4.3.33}$$

where b_i are appropriate beta-function coefficients, and μ is the scale of the process of interest. Since the couplings scale as powers of the age/size of the universe, and therefore meet at 1 at the Planck scale, in first approximation we can assume that, in the effective representation of the physical configuration, couplings run logarithmically with an effective beta-function such that, starting from their "bare" value at the actual $\mathcal{T}^{-1/2}$ scale, they meet at zero at the Planck scale:

$$\alpha_i^{-1} \approx \alpha_i^{-1}|_0 + b_i^{(\mathrm{eff.})} \ln \mu / \mu_0 \,, \tag{4.3.34}$$

with $b_i^{(\mathrm{eff.})}$ such that:

$$b_i^{(\mathrm{eff.})} \ln \mu_0 = \alpha_i^{-1}|_0 \,, \tag{4.3.35}$$

where:

$$\mu_0 \sim \frac{1}{2} \mathcal{T}^{-\frac{1}{2}} \,, \tag{4.3.36}$$

$\mathcal{T}$ being the age of the universe as fixed by the neutron formula 4.3.26. The choice of the square root scale 4.3.36 as the starting scale is dictated by the fact that this is the fundamental scale of matter states, and their interactions.

Let's consider the electromagnetic coupling. The value of α_γ given in section 4.2.1.2 must be considered as a bare value at the scale μ_0. The fine-structure constant, which for us is not really a constant, but just the present-day value of this coupling, will correspond to the value of α_γ run from 4.2.16 at the scale 4.3.36 to the scale μ_γ at which this coupling is experimentally measured. The coupling of the electron to the proton is derived from the wavelength of the atomic spectra, in particular hydrogen. The typical scale is therefore that of the center of mass of the electron and the system of up and down quarks. This is not really the mass scale of the proton itself, which is higher due to the strong interaction of quarks, to which the electron is insensitive.

If we take for μ a multiplicative mean of the electron, up and down quark mass scale as computed in 4.3.10, 4.3.18 and 4.3.19, we obtain as recalculated value of 4.2.16:

$$\alpha_\gamma^{-1} \, : \, \alpha_\gamma^{-1}|_{\mu_0} \; = \; 183.78 \to \alpha_\gamma^{(0)\,-1}|_{m_e} \approx \; 132.9 \pm 0.2 \, . \qquad (4.3.37)$$

The uncertainty is due to the approximation in the choice of the evaluation scale, which is due to our ignorance of the details of the physical system. The result 4.3.37 is definitely closer to the experimental value, nevertheless still quite not right, being out for an amount higher than the error of our approximations. The reason is that the value 4.3.37 has been calculated without taking into account the exchange of the role of up and down quarks, as described in section 4.3.2.3. Modifying the amount of electric charge of a particle results in a modification of the phase space volumes, that, as a consequence of the arguments discussed in section 4.3.2.3, in turn reflects in a different running of the coupling along the mass scales. In order to estimate the order of the correction, we can think that the up-down exchange implies a shift in the mass scales of the order of the mass difference; the relative correction to the coupling should be of the same order of the relative correction of the mass scales:

$$\frac{\Delta \alpha_\gamma^{-1}}{\left(\alpha_\gamma^{(M_0)}\right)^{-1}} \; \sim \; \frac{\Delta m}{M_0} \, , \qquad (4.3.38)$$

where $\Delta m = m_d - m_u$. Lowering the mass of the positively charged quark implies that the scale of the negatively charged lepton, the electron, is now closer to the scale of the quarks with the largest amount of opposite electric charge. The negative charge is spread over a wider range of mass scales (from the electron's scale to the down mass scale, which is now higher than the one of the up quark). As a consequence, the electric interaction gets "screened", or smoothed down, and the coupling consequently lowered. The inverse coupling, α^{-1}, is therefore enhanced. By considering the quark masses 4.3.18 and 4.3.19, we obtain $\Delta\alpha^{-1} \sim 4.12$, and a corrected value of the inverse coupling:

$$\alpha_\gamma^{-1\,(\mathrm{shift})} = \alpha_\gamma^{-1} \, + \, \Delta\alpha_\gamma^{-1} \; \sim \; 137 \pm 0.2 \, . \qquad (4.3.39)$$

It does not make much sense to require a better match with the experimental value of the fine-structure constant (we adopt this terminology by convention, although in our theoretical framework this is not a constant). Indeed, neither the electron mass nor the fine-structure constant are directly measured: they are derived from the value of the Rydberg constant, and the electron magnetic moment. Both these quantities are expressed in terms of α_γ and m_e, so that the relations can be inverted and return the coupling and the mass. However, the quantum gravity corrections, that, as we have seen, affect the value of the mass, affect the value of the coupling too: quantum modifications of the geometry on the small scale reflect in the wavelengths of the observed spectra, which turn therefore out to depend on the bare parameters through modified functions. A detailed evaluation should reconsider all these quantities within this theoretical framework.

We recall that in our framework the electric charge is time-dependent, and 4.3.39 only corresponds to the present-day value of this parameter. The rate of the time variation at present time can be easily derived from the very definition. From 4.2.11 and 4.2.15 we obtain:

$$\frac{1}{\alpha}\frac{d\alpha}{dt} \;=\; \frac{1}{28} \times \frac{47}{45} \times \frac{1}{\mathcal{T}}\,. \tag{4.3.40}$$

In one year, the expected relative variation is therefore of order $\approx 3 \times 10^{-12}$. This is a rather small variation, however not so small when compared to the supposed precision with which α is obtained. Indeed, the most recent measurements give for its inverse a number with precisely 12 digits, a number whose variation could be observed by repeating the measurement at a distance of some years. Since however a fine experimental determination of α depends, through the theoretical framework within which it is derived, on time-varying parameters such as lepton masses etc..., it would not be an easy task to disentangle all these effects to get the "pure α time-variation". This kind of effects can be better detected when expanded on a cosmological scale, as we will discuss in section 5.4.0.1.

4.3.4 The effective strong coupling

The colour coupling α_s is a story apart. Our theoretical framework is rather different from the field theoretical framework in which the experimental value $\alpha_s(M_Z)$ cited in section 4.2.1.4 is obtained. Therefore, it is rather difficult to compare results, especially when they are obtained, as in this case, by interpolating and then running inputs through renormalization equations. In our case, the strong coupling, both in the "strong" and in the dual phase, go to one toward the Planck scale, where all couplings join. The scaling properties are therefore rather different (as are also those of α_W). The interpretation we give of this number is that, as long as the up and down quarks are glued together into a proton or a neutron, although with a varying strength, conceptually they cannot be treated as free particles, not even in an approximated way. This implies that, as long as the proton, or the neutron, hold up, no matter of what is the energy scale at which the proton or the neutron are accelerated, the typical energy to which the colour coupling must be referred to is the one set by the rest energy of the proton, or neutron. Therefore, α_s is not expected to rescale. What does rescale is on the other hand the relative energy scale distance between the center-of-mass energy of the experiment and the rest energy of the neutron, which is related to what we called the "stable" scale, $m_{3/10}$. While close to this energy the stable scale dominates in the phase space when a measurement is performed during an extended time interval (as all experiments are), at higher energies it is possible to observe also the T-dual phase, because its effects are no more so heavily screened by the closeness to the stable scale. This is what in our scenario explains while, although not appearing as free particles, at higher energies it is nevertheless possible to get a glimpse into the fact that proton and neutron are composite particles. If we logarithmically run the value 4.2.22 to the 100 GeV scale, we obtain:

$$\alpha^{\mathrm{T}}_{SU(3)}(M_Z) \approx 0.07 \,. \qquad (4.3.41)$$

A higher value, closer to the one experimentally measured (~ 0.1181), is obtained by rescaling the coupling exponentially, namely, evaluating it at the up-down quark scale, and then letting it run logarithmically up to the 100 GeV scale. This procedure can be justified by considering that a weak-coupling phase of the colour symmetry starts existing at the quark scale, which is therefore probably the scale at which its effective bare value should be evaluated. A discussion of the different behaviour of the strong versus the weak coupling is given in chapter 8. In this case we obtain:

$$\alpha^{\mathrm{T}}_{SU(3)}(M_Z) \approx 0.1 \,. \tag{4.3.42}$$

Although closer to the experimental value than the bare value 4.2.22, there is still a remarkable difference. On the other hand, having to do with a situation very far from that of weakly interacting states, the corrections due to the $m_{3/10}$ scale, as well as the confining phase, must be expected to play a relevant role.

4.3.5 Gauge boson masses

The $SU(2)_W$ symmetry is first broken at the scale at which a mass gap between top and bottom quark is generated. Above this scale, we may consider the symmetry as being (approximately) restored. In section 4.3.2.4 we have seen how these particles acquire a mass. We have also seen that, in the staple of constructions containing a shift in the space-time coordinates, there are certain configurations in which a chiral $SU(2)$ symmetry survives, implying chirality of the weak interaction. This means that the number of configurations in which the momenta of the gauge bosons associated to $SU(2)$ get shifted is lower than the number in which the top quark is shifted. We expect therefore the mass of the W bosons to be lower than the top mass. Since these bosons are "created" through the interaction of top and bottom, we may say that their volume in the phase space is a subvolume of the volume of the top-bottom pair:

$$\mathrm{V}(W) \quad \subset \quad \mathrm{V}(t, b) \,. \tag{4.3.43}$$

The actual volume fraction is given by the part of the volume of the top-bottom space which indeed corresponds to the $SU(2)$ interaction (the top-bottom space can involve also other types of interaction), and is therefore set by the $SU(2)_W$ coupling α_W. The volume at rest is just given by the product of the top and bottom masses, so that the average W mass is given by a coupling-determined fraction of the multiplicative average of the top and bottom mass:

$$M_W^2 \sim \mathcal{O}\left[\alpha_W m_t m_b\right] . \tag{4.3.44}$$

More precisely, the subspace of phase space is shared by the three $SU(2)_W$ bosons that can be exchanged between top and bottom (in physical situations the transition is always among neutral combinations of matter states). Expression 4.3.44 should therefore bear a normalization factor $\frac{1}{3}$. Indeed, the normalization is slightly different, because the neutral boson has a slightly increased phase space. What distinguishes the mass of the Z boson from the one of the $W^{\pm}$ bosons is that the Z boson acquires a "right moving" component: while the charged bosons interact only with a left-handed chiral current, the neutral boson has now a certain amount of coupling with a right-moving current. Since the Z mass is related to the volume it occupies in the phase space, the disagreement between the W and the Z mass is tuned by the strength of $SU(2)_W$ as compared to $U(1)_Z$, the symmetry group associated to the neutral boson. Being the $SU(2)_W$ non-Abelian, its bosons are charged and interact with each other. We can therefore think of Z and W^+ as two particles whose interaction is mediated by W^-. In order to derive the mass of the Z boson, we can therefore consider once again the relation 4.3.44, this time with Z, W^- and W^+ replacing respectively the top and bottom quarks, and the W boson. In this case, we view the process as a transition between W^- and Z, produced by an element of the "group" $SU(2)_W/U(1)_Z$ (more precisely not a group but a coset). The coupling g is now the "coupling" of $SU(2)_W/U(1)_Z$. If we set:

$$\alpha_{SU(2)_W} = \alpha^*_{SU(2)_W/U(1)_Z} \times \alpha_{U(1)_Z} , \tag{4.3.45}$$

we see that, since the $U(1)_Z$ coupling is smaller than the one of the unbroken group, $\alpha^* > 1$. In order to reduce to the ordinary weak coupling the relation 4.3.44 must be S-dualized (this agrees with the fact that, as we saw in section 4.2.1.2, the gauge bosons contribute to the strength of the coupling oppositely to the matter states). Moreover, since we are now considering a transition between bosons instead of fermions, what we obtain is a relation for the square of masses (mass terms are of the type $m^2\phi^2$ instead of $m\psi^2$):

$$\left(\frac{M_Z}{M_W} \right)^2 \approx \alpha^*_{SU(2)_W / U(1)_Z} . \tag{4.3.46}$$

Using the relation 4.3.45, we obtain:

$$M_Z \sim \sqrt{\frac{\alpha_{SU(2)_W}}{\alpha_{U(1)_Z}}} \, M_W . \tag{4.3.47}$$

In order to obtain $\alpha_{U(1)_Z}$ we can proceed as in section 4.2.1.2, this time by determining the fraction with respect to the volume occupied by $SU(2)_W$ at the place of $SU(2)$. This means that the coupling of $U(1)_Z$ should stay to the coupling of $U(1)_\gamma$ in the same ratio as the coupling of $SU(2)_W$ stays to the one of $SU(2)$. Therefore, we expect:

$$\frac{\alpha_{U(1)_Z}}{\alpha_{SU(2)_W}} \approx \frac{\alpha_{U(1)_\gamma}}{\alpha_{SU(2)}} . \tag{4.3.48}$$

Taking all this into account, we can modify expression 4.3.44 and obtain:

$$M^2_W \left(2 + \frac{\alpha_{SU(2)_W}}{\alpha_{U(1)_Z}} \right) \sim \alpha_W m_t m_b . \tag{4.3.49}$$

If in expression 4.3.48 we use the value of α_γ corrected by taking into account the phase space shift around the electroweak scale, and assume in the logarithmic picture a linear running of the inverse corrected coupling from this scale up to the Planck scale, using the values for the top and bottom quark masses, expressions 4.3.18 and 4.3.19, we obtain:

$$M_{W^\pm} \sim 84.0 \, \mathrm{GeV} , \tag{4.3.50}$$

and

$$M_Z \sim 94.9 \, \mathrm{GeV} . \tag{4.3.51}$$

These values are higher than the official ones ($M_W \sim 80.4\,\mathrm{GeV}$, $M_Z \sim 91.1\,\mathrm{GeV}$) but it is not here a matter of obtaining an exact matching with the experimental values. There are too many roughly estimated quantities here. For instance, the scale at which the value of α_W really corresponds: the top scale, the bottom, the W boson scale? A different choice leads to mass values that differ by an amount of the order of our mismatch. But the same could be said about the quark masses, and also those of the unstable leptons. How to correctly evaluate the volume fractions in the phase space? How to correctly account for correction to masses due to the stapling with the stable mass scale $m_{3/10}$? It seems our evaluations tend to overestimate all the masses of the unstable particles heavier than the stable scale (the neutron mass scale, to speak). On the other hand, the Z to W mass ratio is:

$$\frac{M_Z}{M_W} \sim 1.129\,, \tag{4.3.52}$$

a value quite close to the experimental one. This assures that, although several details of the fine evaluation of absolute mass values, and the comparison with the corresponding experimental quantities, are not yet under full control, our procedure is basically correct.

Let us now consider the present-time values of the electromagnetic and the weak coupling, α_γ, α_W, given in 4.2.16 and 4.2.19 ([15]), and the total width of the Z boson, given by 4.3.47: $\alpha_Z = \alpha_W \times (M_W/M_Z)^2$. Their *numerical* relation can approximately be written as:

$$\sqrt{\alpha_\gamma} \;\approx\; \sqrt{\alpha_W}\,\sin\theta\,; \tag{4.3.53}$$

$$\sqrt{\alpha_Z} \;\approx\; \sqrt{\alpha_W}\,\cos\theta\,, \tag{4.3.54}$$

where $\cos^2\theta \approx M_W^2/M_Z^2$. The angle θ can therefore be identified with the Weinberg angle, $\theta \sim \vartheta_W$. Indeed, since the Z boson has a larger width than the W boson only because it has a part of non-chiral

[15] In our theoretical framework, the ratio of these couplings remains the same at any scale.

interaction similar to the one of the photon, these relations say that from an effective point of view we have reproduced the first order of the electroweak gauge sector of the effective action of the Standard Model (except from the Higgs sector, of course: we don't have a Higgs field). The degrees of freedom we have obtained and their interactions can therefore be parametrized in a similar way, namely with interaction terms of the type $g\, J^{\pm}_{\mu} W^{\mp\, \mu}$ and $\frac{g}{\cos\vartheta_W}\left(J^0_{\mu} - \sin^2\vartheta_W J^{e.m.}_{\mu}\right) Z^{\mu}$. The Z^{μ} term precisely says that the Z boson has total width $\alpha^{\rm eff}_Z \sim \frac{g^2}{4\pi\cos^2\vartheta_W}(1-\sin^2\vartheta_W)^2 = \alpha_W\cos^2\vartheta_W$. We stress however that in our case the relation 4.3.53 holds only at the numerical level, it is not a true functional relation. In our theoretical framework the gauge interactions are only an effective first order parametrization of what results from 2.1.16 and 3.1.4.

4.3.6 Mass corrections: the unstable particles

During the finite time interval of an experiment, the stable scale $m_{3/10}$ staples to the fluctuations due to the unstable particles, and corrects them to a lower scale if they are heavier, to a higher scale if they are lighter. Although we are not able to correctly evaluate in detail this phenomenon, we try to provide here an effective parametrization, that reproduces the above described behaviour, leaving a deeper analysis for the future. In the evaluation of the masses of unstable states, we can parametrize the effect of the stapling with a stable mass scale by introducing an *effective* coupling $\alpha_{\rm eff}$, that collects all these effects in the form of an interaction with the stable scale. As it is for the couplings of our theory, here too we assume that the coupling is given by the mass ratio, but, as it enters in the correction of masses at the second order, now the effective relation is:

$$\alpha^2_{\rm eff} \;\sim\; \frac{m_{\rm stable}}{m}\,. \tag{4.3.55}$$

The effective correction should then be:

$$m \;\rightsquigarrow\; m\,(1 \pm \alpha_{\rm eff}) \mp m_{\rm stable}\,, \tag{4.3.56}$$

where the sign of the correction is chosen in this way: for masses higher than the stable scale, the correction has a minus sign, because it lowers the mass, whereas for masses lower than the stable scale it rises the mass. Notice that we don't have a universal coupling, but an effective strength that depends case by case on the mass ratio to the stable scale. The latter is mostly set to be the proton mass scale, as the interaction mainly occurs with this particle. Using this effective parametrization, we can correct the masses to:

$$m_c \;\to\; 1.29\,\mathrm{GeV} \tag{4.3.57}$$
$$m_b \;\to\; 4.09\mathrm{GeV}\,; \tag{4.3.58}$$
$$m_t \;\to\; 173\,\mathrm{GeV}\,. \tag{4.3.59}$$

The official values are $m_c \sim 1.29\,\mathrm{GeV}$, $m_b \sim [4.15 - 4.68]\,\mathrm{GeV}$ and $m_t \sim [172 - 173]\,\mathrm{GeV}$ respectively. In the case of the strange and up/down quarks, this procedure should produce the π and K mass. Indeed, we find:

$$m_{u+d} \;\rightsquigarrow\; m_\pi \;:\; \sim 90\,\mathrm{MeV}\,; \tag{4.3.60}$$
$$m_{s+d} \;\rightsquigarrow\; m_K \;:\; \sim 506\,\mathrm{MeV}\,, \tag{4.3.61}$$

where we have used $m_{\text{meson}} \sim m_{\text{quarks}}\,(1 + \alpha_{\text{eff}}) + m_{\text{quarks}}$. If instead of using the up plus the down quark mass we use twice the down mass, we obtain a pion mass $m_\pi \simeq 122\,\mathrm{MeV}$. These computations are to be taken as a very rough approximation, just an attempt to parametrize the result of a staple of configurations by converting it in terms of effective interaction. For the W bosons, a relation similar to 4.3.55 seems to occur, however to a higher order, in agreement with the fact that these states are not matter states but intermediate states linking matter states:

$$\alpha_{\text{eff}}^{\frac{3}{2}} \;\sim\; \frac{m_{\text{stable}}}{m}\,. \tag{4.3.62}$$

Through this, we obtain:

$$M_W \;\to\; 80\,\mathrm{GeV}\,, \tag{4.3.63}$$

and, from 4.3.52, we obtain also:

$$M_Z \;\to\; 91\,\mathrm{GeV}\,, \tag{4.3.64}$$

values which are in agreement with the official ones. The different power entering the definition of the effective coupling as compared to 4.3.55, namely $3/2$ vs $2 = 4/2$, can be justified as follows. $4 : 3$ is the ratio of the number of degrees of freedom of a massive matter state to that of a massive vector boson of a broken gauge symmetry. The higher is the number of degrees of freedom, the lower is the relative weight of the correction introduced by the superposition with the stable scale, because the higher is the occupation in the phase space of the particle under consideration. In a logarithmic picture, products become sums and ratios differences. The volume V of the particle is the sum of the contribution of the single degrees of freedom. Under the hypothesis that all of them contribute by an amount of the same order, for n degrees of freedom we have therefore, up to a normalization factor, that translates into a common factor at the exponent:

$$n \ln \alpha_{\mathrm{eff}} \sim \Delta \ln V \;\Rightarrow\; \alpha_{\mathrm{eff}}^{n} \sim V_{\mathrm{stable}}/V \;=\; m_{\mathrm{stable}}/m\,. \tag{4.3.65}$$

If from a qualitative point of view these results indicate how indeed a stable scale may affect these masses by rising the lower ones and lowering the higher ones, they must be taken as just an indication. A real computation of these effective values requires a better understanding of the effects of the $m_{3/10}$ scale on the staple of configurations.

4.3.7 The Fermi coupling constant

We are now in a position to make contact with the experimental value of the weak coupling. This is measured through the so-called Fermi coupling constant G_F, a dimensional ($\left[m^{-2}\right]$) parameter defined as the effective coupling of the weak interaction at low transferred momentum [16]:

$$\frac{G_F}{\sqrt{2}} \;=\; \frac{g^2}{8M_W^2} \;=\; \frac{\pi\alpha_W^{\mathrm{eff}}}{2M_W^2}\,. \tag{4.3.66}$$

With α_W^{eff} we indicate here the effective value of the weak coupling, derived as a function of the electromagnetic coupling and the Weinberg

[16] Low means here negligible when compared to the W-boson mass.

angle (or, equivalently, the W to Z boson mass ratio). It is therefore differently defined from the value we have used in order to derive the value of the W boson mass from the top and bottom quark mass, relation 4.3.49. Inserting our results for the W-boson mass, 4.3.63, the Weinberg angle as derived from 4.3.52, and the electromagnetic coupling from 4.3.39, we obtain:

$$G_F|_{M_W} \approx 1.7 \times 10^{-5}\,\mathrm{GeV}^{-2}\,, \tag{4.3.67}$$

a value close to the experimental one ($G_F = 1.166 \times 10^{-5}\,\mathrm{GeV}^{-2}$, see ref. [63]). As it was for the case of the fine-structure constant, once again we are faced with the problem of understanding what is the meaning of a physical quantity, whose value is always related to a certain experimental process at a certain scale. From an experimental point of view the Fermi coupling is obtained by inspecting the pion into muon decay. For our computation we assume that, since both the muon and pion width are affected by the $m_{3/10}$ scale, the "intrinsic" ratio of their widths is set at the level of free quarks, which also set the scale of the effective coupling. If instead of 4.3.63 we use the experimental value of the W mass (80.4 GeV) we obtain a result closer to the one of the literature.

4.4 Flavour mixing and CP violation

As one may expect, in our approach also the mixing of quark flavours in weak decays must be considered in the light of the volume occupied by the various decay channels in the phase space of all possible configurations. The usual classification into families, and the Lagrangian one derives for an effective action, are here justified only by their "statistical" convenience. As a matter of fact, there are no transitions in principle forbidden, but only rare as compared to other ones. The experimental observation that mass eigenstates are not weak-interaction eigenstates is traditionally encoded in a matrix V_{CKM}, the Cabibbo-Kobayashi-Maskawa matrix, which encodes all the information about the "non-diagonal" propagation of elementary particles. It is defined as the matrix which rotates the base of "down" quarks of the $SU(2)$ doublets, allowing to express the current eigenstates in terms of mass eigenstates:

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} . \tag{4.4.1}$$

V_{CKM} accounts for the mixing among different generations, as well as for a CP violating phase. Despite the elegance of the formal treatment, and the intriguing relation between number of quarks and the existence of a phase, from the point of view of the Standard Model the entries of the CKM matrix remain external inputs, chosen to fit experimental data: there seems to be no deep reason why a mixing of quark generations should exist at all, nor why there should be a phase responsible for CP violation. The ordinary theoretical treatment simply provides a parametrization of the quark mixing, for which the number of quark families results to be precisely the minimal one allowing the existence of a phase giving rise to CP violation. In our theoretical framework, the mixing occurs because the phase space of a lighter family can be viewed as a subspace of the phase space of a heavier family: $[3] \supset [2] \supset [1]$. It can be roughly parametrized by a mixing matrix, but the latter must be viewed as just an approximate effective representation of a non-field-theoretical phenomenon. In order to make contact with the Standard Model representation, in the following, we will estimate the entries of this matrix, as they can be computed for an effective action derived within our theoretical framework. However, we will only give the absolute values of the matrix entries, namely the parameters accounting for the amplitude of the non-diagonal decay channels. In our framework, the violation of CP is not the consequence of the existence of a non-reabsorbable phase in a complex CKM matrix, but originates from the general breaking of any kind of symmetry and parity due to the superposition 2.1.16, as a consequence of the implied non-invariance of the time evolution under time-reversal, both at the cosmological and local physics levels.

In our framework neutrino are massive, and therefore can mix and oscillate. Indeed, as a consequence of 2.1.28 they not only can but do necessarily mix and oscillate. Our theoretical framework leads however to a pattern of oscillation that differs in several aspects from the one typically assumed in the Standard Model. We will therefore discuss some cases starting from their experimental detection.

4.4.1 Reproducing the CKM matrix entries

According to our previous discussion, the ratios between entries of the CKM matrix should be of the same order of the ratios of the phase space volumes of the various families, expression 4.1.8. However, a comparison with the experimental results must take into account the conditions under which certain quantities are measured. Needless to say, the involvement of the stable scale $m_{3/10}$ is particularly relevant for the lighter families, for which the experimental energy range falls close to this scale.

In order to make contact with the ordinary description of the mixing mechanism, we must consider that, as it is defined, the CKM matrix is unitary, and collects the information about flavour changing, subtracted of any dependence on masses: in expressions of amplitudes, this dependence is carried by other terms. When translating the entropy-sum driven scenario into the parameters of an effective field theory, one has to consider *how* quantities are measured, namely what kind of

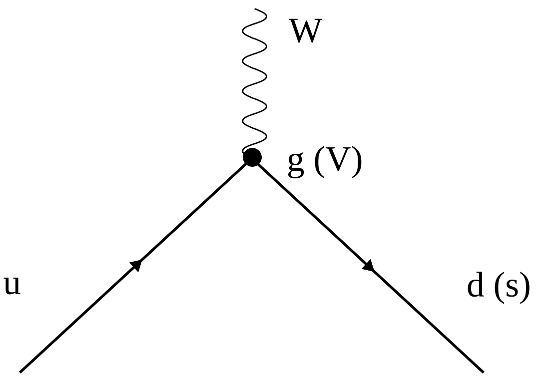


Figure 4.4: Quark mixing corrects interaction vertices: $g \to gV$.

experiment we want to effectively describe. Quark mixing is measured via meson decays into other mesons, and can be parametrized through corrections to the coupling strength of their interaction:

$$g \;\to\; g \times (m_i/m_j)\,, \tag{4.4.2}$$

where i and j indicate the mesons π, K and B.

The CKM matrix elements are of type $V_{\mathrm{up,down}}$ and can be viewed as corrections to the effective coupling of the vertices. This implies that the amplitudes are proportional to:

$$|V_{\mathrm{up}_i,\mathrm{down}_j}|^2 \;\approx\; (m_i/m_j)^2 \tag{4.4.3}$$

where (m_i/m_j) is the ratio of the up (or down) mass of family i to the the up (or down) mass of family j (see figure 4.4) [17]. The CKM matrix entries are generated by unitarity from V_{us}, V_{ub} and V_{cb}. The relation of amplitudes to the elementary degrees of freedom is here complicated by the fact that quarks enter into the amplitude expressions through

[17]Notice that, while the matrix elements relate the "up" of one family to the "bottom" of the other one, the $SU(2)$ symmetry relates bottom to bottom or up to up states.

multiplications and resummations. Strong coupling corrections and the presence of several intermediate decay channels that contribute through a non-direct flavour changing to the overall decay amplitudes correct the ratios we would infer by simply considering bare mass values. The volumes of the events are therefore not so simply related to ratios of bare masses. A better approximation is obtained by using meson masses instead of quark masses. The CKM matrix is built as product of three matrices:

$$\begin{aligned} V_{\mathrm{CKM}} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\vartheta_{sb} & \sin\vartheta_{sb} \\ 0 & -\sin\vartheta_{sb} & \cos\vartheta_{sb} \end{pmatrix} \\ &\times \begin{pmatrix} \cos\vartheta_{db} & 0 & \sin\vartheta_{db} \\ 0 & 1 & 0 \\ -\sin\vartheta_{db} & 0 & \cos\vartheta_{db} \end{pmatrix} \\ &\times \begin{pmatrix} \cos\vartheta_{ds} & \sin\vartheta_{ds} & 0 \\ -\sin\vartheta_{ds} & \cos\vartheta_{ds} & 0 \\ 0 & 0 & 1 \end{pmatrix} ; \end{aligned} \tag{4.4.4}$$

where the entries, derived from quark and meson phase space volumes, correspond to the rotation between second and third, first and third, and first an second family respectively. In order to proceed to an evaluation of the matrix entries, we start therefore by considering the non-diagonal elements $|V_{ts}|$, $|V_{td}|$ and $|V_{us}|$. According to the discussion of section 4.2.1.5, the coefficient relating the third to the second family, should be of the order of $\delta^{\frac{2}{3}}$, the ratio of the volumes of the third to the second family, where δ is the inverse $SU(2)$ coupling, to be evaluated at the charm/bottom scale. It must be remarked that the ratios of volumes determining these three matrix elements relate either the ups or the downs of the two families, not the up of one family to the down of the other. This implies $\delta \sim 100$ and:

$$|V_{ts}| = \sin\vartheta_{sb} \ \sim \ 0.046\,. \tag{4.4.5}$$

In the case of $|V_{us}|$ the phase space volumes are heavily corrected by the closeness to the $m_{3/10}$ scale. For an estimate of these entries, it is better to directly consider the ratio of the K to the pion mass, which already accounts for these corrections:

$$|V_{us}| = \sin\vartheta_{ds} \;\sim\; \frac{m_\pi}{m_K} \;\sim 0.22\,. \tag{4.4.6}$$

The coefficients $|V_{td}|$, accounting for the mixing of the third to the first family, can be inferred as a product of volumes:

$$|V_{td}| = \sin\vartheta_{db} \;\approx\; \delta^{\frac{2}{3}} \times \frac{m_\pi}{m_K} \;\sim\; 0.009\,. \tag{4.4.7}$$

These values are to be compared with those officially reported: $V_{ts} = (40.0 \pm 2.7) \times 10^{-3}$; $V_{us} = 0.2248 \pm 0.0006$ and $V_{td} = (8.2 \pm 0.6) \times 10^{-3}$ respectively (see Ref. [63]). Plugging these values in 4.4.4 we finally obtain:

$$V_{\rm CKM} \;=\; \begin{pmatrix} 0.97549796 & 0.219999569 & 0.00198 \\ -0.219855965 & 0.974447209 & 0.04599991 \\ 0.008190555 & -0.045308133 & 0.998939482 \end{pmatrix}\,, \tag{4.4.8}$$

which parametrizes the quark mixing at the present age of the universe.

4.4.2 Neutrino oscillations

In the case of neutrinos, the detection of the mixing does not occur like in the case of quarks. The rotation among neutrino families does not reflect in corrections to the vertices of a one-loop interaction: the phenomenon we consider involves free propagating neutrinos. Therefore, the probabilities are related in simple way to mass ratios of bare neutrinos. In our scenario, what we can determine is the overall amplitude for the mixing of the muon's to the electron's neutrino (or to the tau neutrino), given by the ratio of volume of the first to the volume of the second neutrino family, namely the mass ratio of the respective neutrinos (or the ratio of the second to the third family respectively). Let us concentrate on the first two families. In the Standard Model approach neutrino oscillation probabilities go typically as:

$$P \;\sim\; \left|\sum U U^\star e^{-i\frac{m^2 L}{2E}}\right|^2\,, \tag{4.4.9}$$

where U are the entries of the Pontecorvo-Maki-Nakagawa-Sakata matrix (PMNS matrix), m the neutrino mass, L the travelled distance and E the average neutrino energy. This expression can be re-written as:

$$P \;\sim\; |UU^{\star}|^2 \sin^2 \left(\frac{\Delta m^2 c^3 L}{4\hbar E} \right) . \tag{4.4.10}$$

The argument of the $\sin^2$ function can be rewritten as:

$$1,27 \times \frac{\Delta m^2 L}{E} \tag{4.4.11}$$

where L is in Km, E is in GeV, and m^2 is in eV^2. Indeed, the $\sin^2$ behaviour is a quite general fact which is a direct consequence of the wave-like propagation, and of being probabilities defined as squares of amplitudes. Therefore, this behaviour does not depend on the specific model through which we implement the mixing, and can be assumed to hold also in our scenario. However, owing to the huge mass difference between muon's and electron's neutrino, in our scenario during one period of the electron neutrino wave the muon neutrino undergoes many oscillation periods. In practice, it just contributes for an averaged effect: the electron neutrino wave projects onto a constant muon neutrino state. The $\sin^2$ argument of expression 4.4.10 reduces therefore to:

$$\approx\; 1,27 \times \frac{m^2_{\nu_e} L}{E} . \tag{4.4.12}$$

What matters in our case is therefore the electron neutrino wave. From the values of the neutrino masses at present time, given in 4.3.2, we can see that for $E \sim \mathcal{O}(1)$ GeV the typical period T is:

$$\frac{m^2_{\nu_e} \mathrm{T}}{\mathrm{E}} \sim 2\pi \;\Longrightarrow\; \mathrm{T} \sim \frac{\pi}{1.27}\, \mathrm{Km} . \tag{4.4.13}$$

The overall value of the coefficient $|UU^{\star}|^2$ for the muon to electron transition, summed over all the internal states, is given by the ratio of the phase-space volumes of the first and second neutrino family, namely by the ratio of the respective masses, normalized to the total amplitude:

$$|UU^\star|^2 \;\approx\; \frac{m_{\nu_e}/m_{\nu_\mu}}{1+m_{\nu_e}/m_{\nu_\mu}}\,. \tag{4.4.14}$$

Averaging over the period of the $\sin^2$ part produces a normalization factor 1/2:

$$\frac{1}{2\pi}\int_{2\pi}\sin^2(x)dx \;=\; \frac{1}{2}\,. \tag{4.4.15}$$

Inserting the value of the neutrino mass ratio, and taking into account the integration over the period, we obtain the average value of the $\nu_\mu \to \nu_e$ mixing, that we indicate as M_{12}:

$$M_{12} \;\equiv\; \langle P_{12}\rangle \;=\; \frac{1}{2}\times\frac{m_{\nu_e}/m_{\nu_\mu}}{1+m_{\nu_e}/m_{\nu_\mu}} \;=\; 0.00342\,. \tag{4.4.16}$$

This allows us to *test* the theory on the experiments. The experimental data we are going to consider are those provided by the Super-Kamiokande [65], MiniBooNE [66] and MicroBooNE [67, 68, 69, 70]. These sources of experimental data are particularly important, because they are less dependent on specific hypotheses and models, like for instance the solar model in the case of solar neutrinos. Moreover, fitting both the Super-Kamiokande and the MiniBooNE data is a challenge for a theory of oscillations, not to speak of the puzzling "disagreement" of MiniBooNE and MicroBooNE data: in the most optimistic scenario the Standard Model prediction lies 5-6 standard deviations away from the MiniBooNE data, a result that the MicroBooNE data seem to put into question, although in a seemingly non-explicable way.

4.4.3 Atmospheric neutrinos

Let us start by considering the detection of atmospheric neutrinos at Super-Kamiokande. This experiment compares the same muon's neutrino beam before and after the travel through earth, thereby getting rid of model-dependent systematic errors on the estimation of the absolute amount of neutrinos. Differently from the usual approach, that assumes the interaction of neutrinos during their travel through the earth to be negligible, in our scenario, owing to the shortness of the oscillation's wavelength (of the order of the kilometer) during their travel muon neutrinos can be assumed to be *in the average* electron's neutrinos by a constant mixing fraction M. This reflects in an increased interaction probability: since stable matter is made of particles of the first family, the interaction with matter of the muon's neutrino is in fact of second order in the weak coupling α_w as compared to the interaction of the electron's neutrino. Therefore, when travelling through matter electron neutrinos have a higher scattering amplitude than pure muon neutrinos. As a consequence, owing to the frequent oscillation, during the travel through matter the muon's neutrino beam decreases through its partial mutation to a more interacting state [18].

Let us call I_{ν_μ} the amount of muon neutrinos which can be *measured* at any time of the neutrino flight. This is proportional to the total amount of neutrinos by a factor that we don't know, and we don't actually need to know. We can write:

$$\frac{\partial I_{\nu_\mu}}{\partial t} \approx -IMA_{\nu_e}\,, \tag{4.4.17}$$

where M is the average amount of mixing over the oscillation period, and A_{ν_e} is the scattering amplitude of the electron's neutrino. Since we are interested in deriving the fraction of remaining neutrinos as compared to the initial amount by comparing the amount of decays before and after travelling through the earth (in other words, since we are not interested in absolute quantities but in relative ones), let us normalize I_{ν_μ} by dividing it by its initial value. I_{ν_μ} will therefore always be lower than one. In order to determine A_{ν_e} we consider that between muon's and electron's neutrino there is an $SU(2)$ rotation among first and second family. We can therefore write:

$$A_{\nu_\mu} = \alpha^2_{SU(2)} \times A_{\nu_e}\,. \tag{4.4.18}$$

[18] Oscillations to the tau neutrino can be ignored here, because they do not significantly contribute to the interaction with matter. We can therefore assume their contribution to the detected events to be basically the same before and after the travel through the hearth, so that it gets systematically subtracted from the experimental data.

where $\alpha_{SU(2)}$ is the strength of the families-rotating $SU(2)$ coupling, the group that determines the mass ratios. Its value must be run to the appropriate energy scale. In our case, we evaluate it at the mean energy scale of the beam we want to consider. In turn, at every time A_{ν_μ} is given by the measured amount of muon neutrinos, namely, I_{ν_μ} itself. We obtain therefore:

$$\frac{\partial I_{\nu_\mu}}{\partial t} \approx -I^2 M \alpha^{-2}_{SU(2)} . \tag{4.4.19}$$

Assuming that the cross sections of the electron's and muon's neutrino scattering remain constant during the path through the earth [19] and are the same as at the point of measurement, we can integrate 4.4.19 to:

$$\frac{1}{I^{out}_{\nu_\mu}} = 1 + \alpha^{-2}_{SU(2)} M \Delta t , \tag{4.4.20}$$

where Δt is the duration of the travel through earth. The neutrino mass is so small that we can consider it to practically travel at almost the speed of light. Therefore,

$$\Delta t \approx 0.0425 \, s . \tag{4.4.21}$$

From 4.4.16 we calculate then:

$$M = 0.00342 . \tag{4.4.22}$$

The inverse of the strength of the $SU(2)$ coupling at energy scale $M_0 = 1/2\sqrt{\mathcal{T}}$ in units of c^2 times the Planck mass M_P was obtained in section 4.2.1.1 to be 4.2.12, that we report here:

$$\alpha^{-1}_{SU(2)} = 147.2 . \tag{4.4.23}$$

In order to run its inverse to the 0.1 GeV scale, we multiply this by the logarithmic fraction of the two energy scales, thereby obtaining:

$$\alpha^{-1}_{SU(2)}|_{E=10^{-20}\, M_P c^2} = \alpha^{-1}_{SU(2)}|_{E=M_0} \times \frac{\log_{10}(10^{-20})}{\log_{10} M_0} \tag{4.4.24}$$

[19] This approximation is justified by the fact that the interaction of the electron's neutrino with matter mostly concerns valence electrons, so that the higher density of earth, five times that of the water, does not play any role.

and:

$$\alpha^{-1}_{SU(2)}|_{E=10^{-18}\,M_P c^2} \;=\; \alpha^{-1}_{SU(2)}|_{E=M_0} \times \frac{\log_{10}(10^{-18})}{\log_{10} M_0} \qquad (4.4.25)$$

for energies of 0.1 and 10 GeV respectively (here we approximate the GeV scale as $\sim 10^{-19}$ times the Planck mass scale). Inserting in 4.4.20 these values we obtain:

$$\left.\frac{1}{I^{out}_{\nu_\mu}}\right|_{\langle E\rangle=\mathcal{O}(0.1\,\mathrm{GeV})} \;\approx\; 2.09\,, \qquad (4.4.26)$$

$$\left.\frac{1}{I^{out}_{\nu_\mu}}\right|_{\langle E\rangle=\mathcal{O}(10\,\mathrm{GeV})} \;\approx\; 1.86\,. \qquad (4.4.27)$$

Both the values 4.4.26 and 4.4.27 are in agreement with the Super-Kamiokande results [65], that also report a higher oscillation rate of neutrino events below, but close to, the GeV energy scale.

4.4.4 The MiniBooNE and MicroBooNE results

Let us now consider the case of neutrinos produced in laboratory. In the usual interpretation, both these data and those of atmospheric neutrinos (as well as those of solar and supernova neutrinos) correspond to measurements made at a different phase of the oscillation. Once parameters such as the mass difference and the PMNS mixing angles are fixed by the other experiments, in order to obtain the prediction for the MiniBooNE experiment it remains only to plug a different energy E and distance L in the same expression 4.4.9. Indeed, the experimental data do not fix the parameters in a unique way, but impose constraints on their values. As a consequence, one speaks rather of a range of predictions. Nevertheless, in the most optimistic case the Standard Model fails to account for the experimental result by several standard deviations (> 4) ([66]): the experimental data show a higher degree of mixing than expected. Several solutions have been proposed to this puzzle, typically sticking on the idea of oscillation between neutrinos of comparable masses, therefore with very long oscillation period. This is a natural assumption if one i) tries to justify neutrino masses within a field-theoretical framework, necessarily based on Higgs mechanism and naturalness of Yukawa couplings, ii) as a consequence explains also the Super-Kamiokande results in terms of single period oscillation. In this case one can try to improve the model by introducing see-saw mechanisms involving non-interacting (sterile) highly massive neutrinos, that would contribute to the oscillation without nevertheless being detected.

Let us now see how things look like in our theoretical scenario. Since we are interested in catching the core of the phenomenon, for simplicity we just consider what could be the overall effect collectively accounting for all the channels, at a reference energy of 1 GeV. While in the case of Super-Kamiokande the period of oscillation is short in comparison to the travelled distance, in the MiniBooNE case the detector is placed at around 1/5th of oscillation's wavelength away from the source [20]. This turns out therefore to correspond to just after the point of maximal rate of increase of the mixing probability, when the $\sin^2$ function attains the value $\sin^2 \sim 0.91$. The MicroBooNE experiment is placed some 70 meters upstream (see for instance [67]), where $\sin^2 \sim 0.75$. Therefore, the MiniBooNE data should in first approximation correspond to a mixing:

$$P(\nu_\mu \to \nu_e) \; = \; 0.91 \times M_{12} \; \approx \; 0.0031 \,. \tag{4.4.28}$$

For energies below 1 GeV we obtain a slightly higher value. For instance, at 900 MeV the period is 20% shorter, and we obtain:

$$P(\nu_\mu \to \nu_e) \; = \; 0.95 \times M_{12} \; \approx \; 0.00326 \,. \tag{4.4.29}$$

This has to be compared with the experimental observation for the neutrino channel, here extrapolated from the data of [66]:

$$\sim \; 0.00323 \pm 0.00014 \,. \tag{4.4.30}$$

[20]For an estimation of the travelled distance, we do not consider the whole distance of the detector from the accelerator's target [71], but the length between the absorber and the neutrino detector, $\sim$ 450m, plus half the detector's length/diameter.

According to [66], the best fit of the Standard Model expectation is instead ~ 0.0026, more than 4σ away from the experimental result. Let us now come to the result of MicroBooNE. In this case, for energies of 1 GeV we obtain:

$$P(\nu_\mu \to \nu_e) \; = \; 0.72 \times M_{12} \; \approx \; 0.0025\,, \tag{4.4.31}$$

and for 900 MeV:

$$P(\nu_\mu \to \nu_e) \; = \; 0.82 \times M_{12} \; \approx \; 0.0028\,. \tag{4.4.32}$$

These calculations are very sensitive to the exact position of the experiment. If the actual center of the experiment is some ten meters more upstream [21], we obtain:

$$P(\nu_\mu \to \nu_e) \; \approx \; 0.0024 \;\; (1\,\mathrm{GeV})\,, \tag{4.4.33}$$

and:

$$P(\nu_\mu \to \nu_e) \; \approx \; 0.0027 \;\; (900\,\mathrm{MeV})\,. \tag{4.4.34}$$

The slope of the typical Standard Model oscillation is not uniquely determined by the experimental data. In any case, roughly speaking the MiniBooNE results can be regressed to the MicroBooNE point by considering that the typical wavelength of Standard Model oscillation models is much larger, at least one order of magnitude larger, than the one of this scenario. This implies that, in the space of the small distance between the two experiments, the MiniBooNE most optimistic mixing, 0.26% according to [66], can be regressed almost like a constant (it would at most decrease by some $2\% - 3\%$ if $\Delta m^2 \approx 0.2 \mathrm{eV}^2$). As a consequence, in the average our prediction for the MicroBooNE experiment, 4.4.31–4.4.34, lies below the upper limit of the values allowed by the Standard Model, whereas the prediction for MicroBooNE lies above. The situation is illustrated in figure 4.5, in which we represent the Standard Model upper limit around the two experiments as a dashed line. Our results are therefore in line with

[21] This correction is perhaps necessary in order to account for the different sizes of the MiniBooNE and MicroBooNE detectors: considering also the different size of the respective detectors, it could be that 85 meters is a better estimate of the distance between the centers of the two experiments.

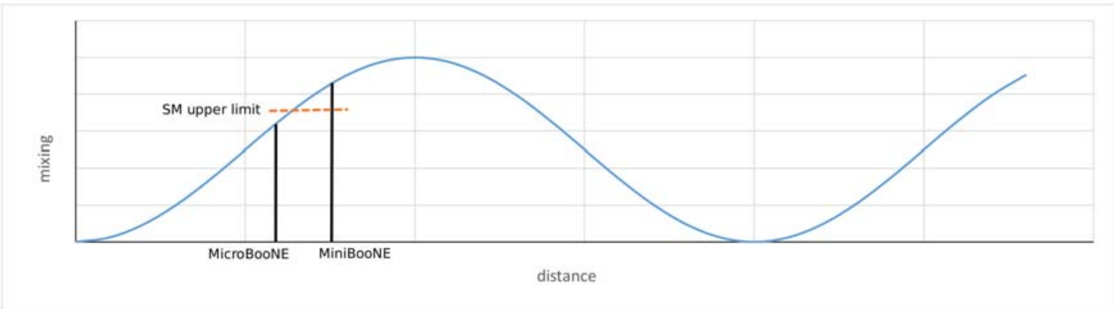


Figure 4.5: The apparent contradiction of the MiniBooNE and MicroBooNE results depends on the architecture of the two detectors, by coincidence placed precisely just before and after the threshold of compatibility with the SM allowed values.

the data reported in Refs. [67, 68, 69, 70], and justify the absence of electron's events excess. According to our analysis, the excess is larger for lower energies, because the wavelength of the oscillation is larger, and the MiniBooNE detector is effectively placed more upstream in the period.

4.4.5 CP violation

In our theoretical framework there is by construction no symmetry under time reversal. Indeed, one can show that, at any time, the staple of configurations gives rise to an observable universe in which all symmetries are broken. However, the time coordinate is something deeply different from the space coordinates. Strictly speaking, there is no "space-time": the concept of "space-time" arises only as an approximation, as part of an effective description of the fundamental scenario in terms of evolving quantum fields in a three-dimensional space. The breaking of the time reversal symmetry is therefore something conceptually different from the breaking of space parity. Nevertheless, in the approximation of relativistic quantum field theory, the general non-time reversal invariance of the overall evolution of the universe reflects, in the microscopic description of physics, into the breaking of the CP symmetry. On the other hand, the parameter of this symmetry breaking cannot be referred to an intrinsic property of a possible mixing matrix of elementary particles, which is in any case just an effective parametrization, only valid in a certain approximation of the fundamental description of physics. In other words, although for practical purposes it is convenient to end up with a description in terms of elementary particles and fields with dynamics determined by the entries of a Lagrangian, the parameters of the effective action are only effective quantities determined time by time, and must be "updated" during the time evolution of the universe. In this perspective, CP violation does not originate from quark mixing terms in a Lagrangian, although the latter can be a useful parametrization for this phenomenon.

4.4.6 Time reversal asymmetry in the phase space of particles

By definition, in this theoretical framework physical amplitudes are not computed out of the ingredients of a Lagrangian formulation at the base of the physical description: they are a consequence of the entropic principle ruling 2.1.16. As a consequence, any Lagrangian description must be viewed as just an effective approximation. The fundamental objects ruling the time evolution in microscopic phenomena are therefore no more the matter degrees of freedom described as asymptotic states that come into interaction through terms of an effective action, and therefore their propagators. Scattering amplitudes are no more fundamentally associated to vertex operators of string theory either. The objects to be considered are now the phase space volumes. This is a different conception of physical evolution, which implies a new way of dealing with dynamics.

Since decay amplitudes (or, in general, interaction amplitudes) correspond to the volumes occupied by the corresponding processes in the appropriate phase space, we must expect that also the amount of violation of time reversal in the weak decays does correspond to an asymmetry in the volume of the phase space for a decay process and its CP-mirror, which reflects the general lack of time-reversal symmetry of the theory at the macroscopic level.

The volume of a process in the phase space depends on several parameters, such as the strength of the coupling, the type of interaction channels, initial and final momentum etc... Typical field-theory interaction amplitudes are then integrated over the range of space momenta. In our case, the full phase-space weight depends also on the energy-momentum four-volume factor $dE\, d^3p$, that accounts for the proper volume of the state under consideration. This latter one is precisely the part of interest for the evaluation of the time asymmetry, because it accounts for the volume occupied by the state in the energy-momentum phase space, which is conjugate to space-time, and, differently from the internal factors accounting for the strength of coupling etc., not only it depends on time, as any parameter in this theoretical framework, but it is very sensitive to the time arrow, and to operations performed on the space-time.

In order to evaluate this term, we must consider that a particle is an energy packet which has a non-vanishing extension both in time and space. The energy spread is of the order of the mass: $dE \sim m$. Quantum-relativistic arguments, together with the observation that in this scenario no particle can have a radius smaller than the Planck length size, imply that also the momentum spread, in each of the three space directions, must be, in appropriately converted units, of the order of the mass. All this implies $dE\, d^3p \sim m^4$. In the case of the initial state, this is therefore of order m_i^4. For the final state, the decay product, it is m_f^4. At the decay point the phase space volume is increased by the added possibility of interpreting the energy packet also in terms of the final states. The overall four-volume of the process is therefore the sum of the volume of the initial and of the final states: $\sim (m_i^4 + m_f^4)$. Once all the factors accounting for the detail of the interaction are factored out, and the proper volume is normalized to the volume of the initial state, the decay amplitude can be written as:

$$N \propto \left[1 + \left(\frac{m_f}{m_i} \right)^4 \right] . \tag{4.4.35}$$

Let us consider now the decay into the CP-conjugate of the final state. In order to understand the behaviour in this case we must go back to the early interpretation of anti-matter as negative energy matter. In order to compute the variation in the phase space volume due to time reversal, consider that producing an anti-particle is like creating a “hole” at the place of a particle. Therefore, the volume of the produced particle will not be added to, but subtracted from the initial volume as in order to create an energy-momentum hole:

$$\overline{N} \;\propto\; \left[1 - \left(\frac{m_f}{m_i} \right)^4 \right] . \tag{4.4.36}$$

Put in other words, since we are going “backwards in time”, we are destroying volume in the energy-momentum phase-space. We remark that we are generating here a net difference between the two processes, namely decay into a state and decay "backwards in time" to the conjugate state, because in our scenario *every* process occurs during a finite amount of time, during which the history of the universe goes by, entropy increases, and all the weights, and phase-space volumes, change with time. This is not the case of a theory, such as ordinary field theory, in which the fundamental description is time-conjugation invariant, and parity breaking is introduced "ad hoc" by appropriate terms.

From N and $\overline{N}$ we obtain the CP-asymmetry $\mathcal{A}_{CP}$ as $(N-\overline{N})/(N+\overline{N})$:

$$\mathcal{A}_{CP} \sim \left(\frac{m_f}{m_i}\right)^4 . \tag{4.4.37}$$

This approach is somehow "inclusive", not sensitive to distinctions between direct and indirect CP violation. Moreover, it leads to the same asymmetry for decay into neutral or into charged particles. Differences between decay channels can only be revealed by a more detailed evaluation of the phase space volumes at play.

4.4.7 CP violation in meson decays

Let us now test our approach on concrete examples. The first case in which historically CP violation has shown up is the neutral K-mesons system. The K meson is composed by an s and an (anti-) d quark, and mostly decays into pions (one pion plus leptons, or also into more pions at once). The masses of the quarks involved in the transition that characterizes the process, namely $s \to d$, are much lower than the masses of the corresponding mesons, K and π. This means that strong corrections are at work. Indeed, as discussed in section 4.3.6, the effective experimental value of these masses is "perturbed" by the "stable" mass scale of the universe, roughly corresponding to the neutron mass. This correction is due to the fact that experimental values are obtained as average results of events that occur during a certain interval of time, where a stable mass has a relatively higher probability of being detected than a short-life one. In the specific case of the computation of the phase space volumes in the purpose of deriving the size of the

CP violation effect, we can keep into account these effects by using in the expression 4.4.37 for the mass of the incoming particle (m_i) m_K rather than m_s and, for m_f, m_π instead of m_d. The possible presence of other pions as decay products does not affect this computation, because, owing to the factorization properties of the phase space, in first approximation the contribution to the phase space volumes of other particles produced in the decay can be neglected: they can be treated as "spectators". m_i and m_f stay therefore for the mass of the initial and the final meson involved in the quark decay. Inserting the values of $m_{K_0} \approx 497.6\,\mathrm{MeV}$ and $m_\pi \approx 134.98\,\mathrm{MeV}$ we obtain:

$$\mathcal{A}_{CP}^{(K)} \sim 5.4 \times 10^{-3}\,, \tag{4.4.38}$$

to be compared with the asymmetries obtained from experimental measurements (see ref. [63] for a state-of-the-art overview):

$$A_L = \frac{\Gamma(K_L^0 \to \pi^- \ell^+ \nu) - \Gamma(K_L^0 \to \pi^+ \ell^- \nu)}{\Gamma(K_L^0 \to \pi^- \ell^+ \nu) + \Gamma(K_L^0 \to \pi^+ \ell^- \nu)} = (3.32 \pm 0.06) \times 10^{-3}\,, \tag{4.4.39}$$

and similar ones (owing to the high degree of model-dependent elaboration of experimental data, these results are in general to be considered as an indication of the order of magnitude of the effect more than as highly precise estimates). In the case of the D mesons, we have a transition $c \to s$ for the decay $D \to K\pi$, where as before the pion can be treated as a spectator. In this case, the mass of the charm quark is slightly above that of the neutron, and we should expect it to be less affected by strong corrections. Indeed, the D mass is not so different from the mass of the c quark. If we insert in 4.4.37 as initial mass the quark c mass ($\sim$ 1.3 GeV), and as final mass the K meson mass ($\sim$ 498 MeV), we obtain:

$$-\left(\frac{m_K}{m_c}\right)^4 \sim -2.2\,\%\,. \tag{4.4.40}$$

If instead we use the D meson mass (1864.9 MeV) we obtain:

$$-\left(\frac{m_K}{m_D}\right)^4 \sim -0.508\,\%\,. \tag{4.4.41}$$

With an "average" mass, $\langle m \rangle = (m_D + m_c)/2$, we would have:

$$-\left(\frac{m_K}{\langle m \rangle}\right)^4 \sim -0.9\,\%\,. \tag{4.4.42}$$

In 2011 [72] the analysis of experimental data produced an average asymmetry of about $(-0.832\pm0.033)\%$ (see also ref. [73]), in agreement with this estimate. Subsequent revisions in the light of an increased amount of collected experimental data seem to have reduced by about one order of magnitude this value [74, 75]. However, it is difficult to derive final conclusions, because the refinement in the analysis strongly depends on the Standard Model theoretical scheme [63].

A third system in which CP violation plays an important role are the B mesons. In order to give a rough estimate of the order of magnitude of the effect we expect in our theoretical framework, we may consider an average within a range starting from the decay $B \to J/\psi$, based on a transition $b \to c$, and therefore expected to be of order $-(m_c/m_b)^4 \sim -7.6\times10^{-3}$, passing through the channel $B \to K$, for which we better consider the K mass instead of that of the quark s, $-(m_K/m_b)^4 \sim -(498\,\mathrm{MeV}/4400\,\mathrm{MeV})^4 \sim -1.7\times10^{-4}$, to arrive to the semileptonic decay $B \to \ell\ldots$, which gives an almost negligible asymmetry (for instance, for the $B \to \mu$ decay, we have $-(106\,\mathrm{MeV}/4400\,\mathrm{MeV})^4 \sim -3.4\times10^{-7}$). Owing to the high degree of uncertainty, and to the strong dependence of the latter on theoretical assumptions related to the choice of the model to be tested, a comparison with experimental results is difficult, and rather questionable. At present, values of order $\sim 10^{-5}$ are not excluded [63].

Notice the change of sign in the asymmetry between decays from quarks of the second and third family into quarks of the first family (up, down) and decays that involve only quarks of the second and third family. In our theoretical framework, we see this as a consequence of the fact that, as discussed in section 4.3.2.3, owing to the up-down flip in the first family, as a matter of fact the up quark behaves like an anti-down, and a down quark like an anti-up. There seems therefore to be a flip in the effective time arrow between the first and the other two families.

4.4.8 CP violation in neutron decays: the baryon asymmetry

In our scenario there is a priori no condition preventing the occurrence of baryon number violating decays. Similarly to what happens for the condition of three-dimensionality of space-time, also a situation in which there is no baryon number violating vertex, like in the Standard Model, is here recovered only statistically, being the baryon number violating process very rare in the phase space. If we consider a neutron beta decay into proton plus electron and neutrino we find that its phase space volume is much larger than that of the CP-conjugate, baryon number violating decay channel:

$$A_{CP} = \frac{m_{\mathrm{p}}^4}{m_{\mathrm{n}}^4} \sim 0.995 \,. \tag{4.4.43}$$

As one can expect, also in our scenario baryons can be produced out of non-baryonic states through baryon-antibaryon pair production, followed by asymmetric decay, with preference for one of the two CP-conjugate states. Indeed, in the universe one observes a baryon to photon ratio η [76]:

$$\eta = \frac{n_B}{n_\gamma} = (5.5 \pm 0.5) \times 10^{-10} \,, \tag{4.4.44}$$

which can be interpreted as the result of the progressive annihilation of protons against anti-protons during the phase of cooling down of the universe, namely, before the average temperature of photons fell down below the mass-threshold for the proton-antiproton pair production, $T_\gamma < 2m_{\mathrm{p}}$ [22]. In this interpretation, the present value of n_B/n_γ should be what remains of the asymmetry $(n_B - n_{\bar{B}})/n_\gamma$. The Kobayashi-Maskawa mechanism doesn't allow to account for such a high value of the asymmetry as the one which is observed. In our case, the size of CP violation effect depends on time (the age of the universe), and at earlier times it was stronger due to the fact that masses were (relatively) closer to each other. Namely, the absolute value of the difference of masses was larger, because all of them were closer to the Planck scale,

[22] For an introduction see for instance [77].

but the ratio of mass differences to their absolute value was lower. Therefore, from expressions 4.4.43, 4.3.26, 4.2.31, 4.2.32 and 4.2.11 one can see that the amount of CP violation was higher. However, in our scenario also the evolution of the universe occurs in a different way. There is certainly a cooling down, but this is driven by the temperature of the universe as a black hole (see [16]), with temperature $T \sim 1/\mathcal{T}$. The energy densities of matter and radiation are always of the same order, $\rho_{m,r} \sim 1/\mathcal{T}^2$, therefore there is no phase in which there is a sea of photons predominantly with an energy higher than that of matter: the mean energies of photons and matter scale almost in the same way along the history of the universe [17]. In our scenario, the photon abundance, or equivalently the baryon asymmetry, does not come from the pre-history of the universe, but reflects instead a "stationary condition", as we now explain. Let us consider the neutron beta-decay. According to 4.4.43 one would think that from neutrons only protons are produced, and almost no anti-protons. However, the process of proton (or antiproton) production through neutron decay doesn't go on till the complete disappearance of the neutrons. The reason is that the decay products of the neutron, namely the proton, the electron, and the neutrino, are all end-products, which cannot further decay because they are already the particles of minimal mass, at the end of the decay chain. They can instead easily recombine to reproduce the neutron, so that, apart from some unstable isotopes, neutron and proton are found in nature basically in equal number. Owing to this "equilibrium" condition, with good approximation we may think that all the protons existing in the universe come from neutron decays, and that the baryon asymmetry should be computed from the properties of the neutron decay. However, expression 4.4.43 is of no help in deriving the amount of protons (antiprotons) effectively produced, and says nothing about the number of photons one eventually produces as the result of proton-antiproton annihilation. In order to derive the CP asymmetry in the neutron/proton system through an analysis of the volumes of the phase space we must take into account the fact that, unlike the decays considered in the previous section, here we have a process at equilibrium. That means, there is no net change in the volume of the phase space. One starts with a neutron/proton system and ends up again with a neutron/proton system. There is nevertheless a transition, involving the passage from up to down quarks and vice-versa, but this has to be treated as a fluctuation. It can be viewed as a sort of oscillation of the system p, n, e, ν:

$$(p, n, e, \nu) \leftrightarrow (\bar{p}, \bar{n}, e^+, \bar{\nu}) \,. \tag{4.4.45}$$

Consider the transition neutron-proton. There are three quarks involved, namely (u, d, d), which go into (u, u, d). It would seem that, as net change, we just have the decay $d \to u$. However, from the point of view of the phase space this is not so simple. Owing to the fact that, unlike the mesons, neutron and proton are made of three quarks, and therefore are $SU(3)$ singlets in which the colour symmetry mixes up degrees of freedom of all the three quarks, in the transition from neutron to proton all the three quarks are involved, in something like: $u \to d$, $d \to u$, $d \to u$ [23]. During this transition one physically generates a fluctuation in the volume of the phase space corresponding to a mass fluctuation of order $\Delta m = 3\Delta m_{d \to u}$. For what matters the CP violation the volume of the neutron does not count, and the only asymmetry in the phase space is given by the transition $0 \to 0 \pm \Delta m$, where Δm is measured in units of the neutron mass. In order to take into account the renormalization due to the strong corrections, for $\Delta m_{d \to u}$ we don't take the bare quark mass difference, but the neutron-proton mass difference. The so computed CP asymmetry should correspond to one-half of the expression 4.4.44, because for any pair of proton/antiproton which annihilate one produces two photons [24], and we can in first approximation neglect the photons produced by electron-positron and neutrino-antineutrino annihilation (the latter obtained through the intermediate production of a neutral boson), because in general of much lower energy. We obtain therefore:

$$\left[\frac{3(m_{\mathrm{n}} - m_{\mathrm{p}})}{m_{\mathrm{n}}} \right]^4 = \frac{n_B - n_{\bar{B}}}{2\, n_\gamma} \,. \tag{4.4.46}$$

[23] Mesons are instead of type $q\bar{q}$, for which $SU(3)$ singlets are built up diagonally.

[24] Pair annihilation produces a double photon due to momentum conservation (there cannot be a photon with zero momentum).

Inserting the current mass values, we obtain:

$$\left[\frac{3(m_{\rm n} - m_{\rm p})}{m_{\rm n}} \right]^4 = 2.87 \times 10^{-10} \,, \tag{4.4.47}$$

and therefore:

$$\eta_{\rm predicted} \sim 5.74 \times 10^{-10} \,. \tag{4.4.48}$$

Notice that this computation does not rely on the details of the various (virtual) channels, because, like in the CP violating decays, it considers only the net fluctuation between initial and final state. Therefore, the value we obtain in this way in principle accounts for the contribution of all the various virtual channels through which this transition may be figured out to occur. This value is a ratio of two mass scales which have almost the same time-dependence. Therefore, it has almost no time-dependence, and we expect it to approximately correspond to the value 4.4.44, derived in ref. [76] from nucleosynthesis constraints.

As we discussed in section 4.3, in this theoretical framework the colour force is strongly coupled in the strict sense, i.e., the coupling strength is larger than 1 and the colour degrees of freedom are confined to singlets. There are no gluons at all. However, as we saw, although statistically suppressed, in the phase space a certain amount of S-dual phase is also present. This is the part responsible for the fact that, under certain conditions, it is possible to inspect the quark structure. The strong CP problem has here to be addressed in the light of these considerations: the confining part is unaffected, because it cannot be written as gauge theory at all, but the S-dual phase is in principle sensitive to CP violation. This justifies why strong CP violation is very suppressed. The amount of breaking of S-duality substitutes here a suppression mechanism such as for instance the Peccei-Quinn symmetry. From this point of view, it is quite likely that this mechanism, and the related axion fields, suffer the same fate as the Higgs mechanism and field.

4.5 Partial S-duality in the electromagnetic interaction

As discussed in section 4.2.1.4, S-duality is not completely broken, and in principle allows for the existence of situations in which S-dual aspects show up and can be detected. In section 4.3.4 we have seen how this occurs for the quark colour force (presence of the strong and weak coupling phase of α_s). In this section, we discuss in detail the case of the electromagnetic coupling, α_γ (in the case of the weak coupling α_W there is no such a kind of phenomenon, because the $SU(2)_{\mathrm{L}}$ symmetry is broken). We will see that, in this way, we can account for the 125 GeV resonance detected at LHC, usually interpreted as a Higgs signal, and for other two lines in the photon spectrum, detected by analyzing cosmic radiation collected by telescopes (Fermi-LAT), at ~ 111 GeV and ~ 130 GeV.

The sum 2.1.16 accounts for the whole universe. Space-time is not factored out: the effective geometry resulting from the stapling of geometries varies from point to point. It is only in the string representation that, for technical reasons, in order to make possible a perturbative construction, gravity is basically decoupled, and the space-time appears as factored out. This produces a description in which the microscopic world appears to be the same at any point of space-time, allowing to investigate the spectrum of elementary states, to be used as the building blocks of an interacting world. In the language of 2.1.16, making experimental measurements means looking at certain selected regions of space-time. In this way, we "filter" the configurations that contribute in a relevant way to the particular effective geometry we are investigating, and therefore to the mean value of observables. When we decide to look at a particular phenomenon under certain conditions, for instance a scattering of particles at a certain energy scale, we effectively operate a selection in the staple of all possible configurations. For instance, we may look at certain physical systems under conditions that favour the appearance of S-dual aspects of the one or the other coupling [25]. We want to see what are

[25] A selection of this kind is always implied by the choice of the scale at which to look at certain phenomena, which determines whether we are in a regime of classical

the conditions that favour the appearance of a strong coupling phase of the electromagnetic interaction. Let us consider the collision of two particles of mass m_1 whose interaction corresponds to a gauge symmetry group G, with coupling $g < 1$ (weak coupling). In a S-dual phase ($g \to 1/g$), at a center-of-mass energy $E_{\text{c.o.m.}} \sim 2 \times (1/g^2) \times m_1$, at the point of collision the two particles behave like one single particle of mass $m_2 = m_1/g^2$. This implies that, at this center-of-mass energy, we should expect an increase of the scattering amplitude, due to the extra channels that precisely at this energy concur to the process: besides those of the two particles with mass m_1, also those of the particle with mass m_2. On the other hand, the particle with mass m_2 can be viewed as derived from the particles with mass m_1 by "eating" the degrees of freedom of one group factor G, which now contribute to the *internal* symmetry of the particle with mass m_2, thereby increasing its mass. In the freezing of the degrees of freedom due to the strong coupling, the eaten volume is in fact precisely proportional to the S-dual of the coupling in the weak coupling regime, $1/g^2 \sim 1/\alpha$ [26].

4.5.1 Ratios of volumes in the phase space

Let us focus our attention on the electromagnetically charged particle-antiparticle pairs. We want to see more in detail under what conditions it is possible to produce an effective strongly coupled equivalent state, leading to an increase of the scattering cross section. In order for this to occur, it is necessary that the gauge degrees of freedom of a pair of independent particles can be interpreted as collapsing, due to the strong coupling, to a configuration in which there is no gauge group at all (frozen gauge degrees of freedom): this configuration is therefore

or of quantum mechanics. In practice, it is like selecting the value of $\hbar$ in the Feynman path integral.

[26] In some sense, the tower of elementary particles, whose masses are at a distance set by the inverse of the $SU(2)$ coupling, can be viewed as a hierarchy of massive states obtained by eating the degrees of freedom of lighter ones. The hierarchy of elementary particles could therefore be considered as as a realization of this phenomenon in the case of the $SU(2)$ symmetry, which would therefore give the weak interactions in the weak coupling phase, and the set of massive states in the strong coupling phase...

equivalent to the one of an electrically neutral state. The gauge degrees of freedom contribute to the volume occupied in the phase space by the two-particles configuration by a factor $V_{(\text{particle}\,1)} \times V_{(\text{particle}\,2)}$, which in this case is reduced to 1 (just one possible configuration, no enhancement proportional to the size of the orbit of a symmetry). The only possibility for this to occur is that the scattering does involve hadrons, either as intermediate states, e.g. when in the collision of a e^+e^- (or other lepton-antilepton) pair one creates a $p\bar{p}$ (or heavier hadron) pair, or as colliding particles, e.g. a collision of a $p\bar{p}$ pair, where one creates either lepton pairs (e^+e^-, $\mu^+\mu^-$ or $\tau^+\tau^-$) or even other hadron pairs. Incoming and intermediately produced particles must then couple in order to form electrically neutral compounds that contain *more than two spinors*. For instance, if we let to collide a $p_{\text{in}}\bar{p}_{\text{in}}$ pair, it must be possible to create an e^+e^- pair, which forms bound states of the type $[p_{\text{in}}e^-]$ (and/or their charge-conjugates), as shown in figures 4.9 and 4.10. The reason is the following. In a lepton-antilepton pair the electromagnetic group acts in opposite way on the two states of the pair, by a transformation depending on a parameter β: $\bar{\psi}\psi \to \mathrm{e}^{-\mathrm{i}q\beta}\,\bar{\psi}\psi\,\mathrm{e}^{\mathrm{i}q\beta}$, where q is the actual value of the electric charge Q. The non-effectiveness of the overall transformation is attained as the result of an exact point-wise cancellation all over along the orbit of β, a situation effectively equivalent to having zero electric charge, like in a neutral state: $\phi_{Q=0} \to \mathrm{e}^{\mathrm{i}0\beta}\phi_{Q=0}$. The volume of the orbit, $V(\beta)$, is the span of all the values of the parameter β, and is clearly the same for any value of the electric charge Q, so that in both the cases it is the same: $V_q(\beta) = V(\beta)$, $\forall q$. Therefore, when the particle's compound can be considered equivalent to just one neutral particle, there is no change in the volume of the orbit in going to the strong coupling [27]: the volume of the group passes from being $V_{Q=q}(\beta)$

[27] This theoretical setup is originally defined on the discrete space, and determining volumes is in principle a simple thing. The volume of a discrete group is simply the number of its elements. However, the contact point with the physics we experience, and test, occurs in the limit to the continuum, where we recover the familiar concepts of field theory, and gauge groups. In this limit, things are no more so obvious. But, since we are eventually interested in ratios of volumes, we don't really need absolute values of group volumes, but relative ones. In this case, it is still natural to think of the volumes of compact Lie groups as given

for the lepton-antilepton pair to $V_{Q=0}(\beta)$ for the neutral bound state. On the contrary, in the case of the lepton-hadron compound, like the $[p_{\rm in}e^-]$ pair, the effective charge cancellation occurs through a sum of Lie-group parameters $\beta_{\rm i}$:

$$\beta_{u_1} + \beta_{u_2} + \beta_d = -\beta_e \,. \tag{4.5.1}$$

There are therefore two more free parameters than in the case of the lepton-antilepton pair. As compared to the case of a single neutral particle, the set of two particles has two group volume factors more. At the energy of effective strong coupling the same volume of occupation in the phase space can be equivalently viewed as corresponding either to two particles, pointwise paired (p, e), or to a configuration with a single neutral particle (the strongly bound $[p - e]$ compound). In this second case, the volume occupied in the phase space has to be interpreted as entirely due to rest energy (= mass) of the neutral particle. The mass gap between the neutral particle and the two single particles corresponds to the volume of the missing electromagnetic symmetry group, namely the *product of the volumes*, corresponding to two $\beta_{\rm i}$ parameters: $M_{[p-e]}/M_{p+e} \sim [V(\beta)]^2$. In order to see the relation to the coupling g, we must consider that, as it is defined on the Lie algebra, the coupling g works as unit of measure of the values the Lie parameter (in gauge theory promoted to local field) can assume along a period of the orbit:

$$g \times \text{Volume} \simeq 2\pi \,. \tag{4.5.2}$$

The rest energies of the two configurations stay therefore in a ratio given by:

$$\frac{M_{p+e}}{M_{[p-e]}} \simeq g^2 \,. \tag{4.5.3}$$

In both these expressions we omitted the exact normalization of the coupling. Indeed, this is fixed by requiring that, by definition, the

by the volume of the space of their parameters, and therefore determine ratios of volumes as the ratio of the number of generators of the Lie algebra (the ratio of dimensions).

ratio of the phase space amplitudes is precisely the coupling $\alpha \stackrel{\mathrm{def}}{\equiv} \frac{g^2}{4\pi}$ (see section 4.2). Relation 4.5.3 can be expressed as:

$$M_{[p-e]} \sim \alpha^{-1} M_{p+e}\,. \tag{4.5.4}$$

This situation has to be compared with a typical expression of binding energy in the weak coupling regime. For instance, in an hydrogen atom the electronic energy levels, which refer to standing waves and are derived from the Coulomb potential, therefore a second-order effect in powers of the coupling, are proportional to $[m_e]\alpha^2$. Here we have instead a first order effect in the S-dual of the coupling: α^{-1} vs. α^2.

4.5.2 The 125 GeV resonance at LHC

Let us now consider the dynamics of a particle-antiparticle scattering. As seen from a geometric point of view, what we have is a cluster of energy around the scattering point. When the amount of energy allows the interpretation of the cluster not only as a set of weakly interacting elementary particles, but *also* as bound state of strongly coupled particles, the combinatorial possibilities increase. This implies a larger volume in the phase space (i.e. a larger volume of the combinatorial group of the distribution of energy). In our theoretical framework, this translates into a relation similar to 4.5.3, this time referred to the full effective coupling of the interaction, $M_i/M_f \approx g^2_{\mathrm{eff}}$, the ratio of the whole weight of the initial configuration to the weight of the final scattering products. Adding new combinatorial possibilities to the initial configuration before the scattering, i.e. increasing M_i, leads to an increase of the effective coupling, and therefore of the scattering amplitude.

If we want to look at the details of what is going on, and recover the familiar description in terms of elementary particles and their interactions, we must leave the phase space and look at the time evolution of the process. In the phase space, at any given time geometries are summed at that fixed time to contribute to the average geometry at that time of the physical evolution. Scattering amplitudes are instead obtained by summing up scattering events along a certain interval of time, corresponding to the duration of the experiment [28]. It turns out that the energy $E = \alpha_{\gamma}^{-1} M_{p+e}$ is a critical energy, at which new possibilities of realizing the scattering open up. Consider a proton-antiproton scattering. At the critical energy, the scattering amplitude receives comparable contributions not only from a first order process like the direct $p\bar{p} \to \gamma\gamma$ scattering: also channels which in an ordinary perturbative expansion over the value of the coupling would be suppressed contribute in this case with comparable strength. This occurs only at the critical energy, when an interpretation in terms of strong coupling opens up: out of this point, these channels are ordinary higher-order processes, and are therefore suppressed. The additional channels involve the creation of a lepton-antilepton pair (e.g. electron-positron pair). In the usual perturbative approach, represented by Feynman diagrams in terms of interactions and propagators of free fields and particles, these are second- and third-order processes, as illustrated in figures 4.7 and 4.8. However, at the critical energy of effective strong coupling there is no g^2 and g^4 suppression, because the proton-electron pairs are strongly bound into one state. As a consequence, there is no gauge symmetry with coupling g mediating the interaction among particles within the bound state (see figures 4.9 and 4.10). Therefore, these decays into pairs of photons sum up with a strength comparable to the one of the first-order, direct $\to \gamma\gamma$ process. Above the critical energy, owing to mismatching momentum account we can no more interpret the energy cluster as the bound state plus the other free particles. The reason is the following. At the threshold the total

[28] A conceptual difference between this approach and the scattering probabilities as they are defined in quantum mechanics needs here to be pointed out: in the traditional approach to quantum mechanics, probabilities are defined at any instant of time, and are compared with experiments performed along a time interval. Here, speaking in terms of probabilities is not much appropriate: in this scenario physics is neither deterministic nor probabilistic. It is rather "determined", as the result of an infinite number of contributing terms. It is precisely the infinity of contributions, and the impossibility of interpreting all of them in terms of "classical" geometries, what forcedly leads to an interpretation in terms of probabilities. In this scenario, speaking in terms of probabilities is considered a (unavoidable) conceptual artifact, allowing to encode, and predict, experimental results, because a parametrization in terms of the usual concepts of particles, masses, couplings, is only allowed above a certain scale (of space, time, energy).

energy equals the sum of the masses of the involved particles. Above this energy, there must be also some momentum. But the existence of a bound state, in this case a (pe) bound state, implies that p, e, p^- and e^+ have *all* the *same* speed, whereas the first two are paired to a higher mass state. This is incompatible with energy-momentum conservation; above the critical energy the bound state channels are suppressed once again, as they were below the threshold.

To summarize, the scattering amplitude has a peak centered around the critical energy of effective strong coupling, characterized by an excess of typically leptonic decays, ($\ell\bar{\ell} \to \gamma\gamma$). These processes are illustrated in figures 4.6, 4.7, 4.8, 4.9 and 4.10. Picture 4.6 shows the basic, first order $p\bar{p} \to \gamma\gamma$ process, which is going to be reinforced at the critical energy by the contributions illustrated in figures 4.7–4.10. They show scattering channels which, from a field theory point of view, are of higher order in the coupling α_γ. They are therefore suppressed, except at the critical energy, where one can interpret the intermediate virtual components as forming a strongly coupled compound, accompanied by the disappearance of the gauge symmetry associated to their interaction [29]. At this point, and only at this point, they are no more of higher order (i.e., no more suppressed). The phenomenon we have described is not a property of just the $p\bar{p}$ scattering. The enhancement of the cross section occurs, under the same conditions, and at the same critical energy, also if in the diagrams 4.6–4.10 one exchanges proton and electron: in lepton-antilepton pair scattering, via creation of a proton-antiproton pair, in which one or both the hadrons couple in a strong way to the electron and/or the positron. Analogous considerations can be done with the charged leptons μ and τ at the place of the electron. The peaks of cross section they produce by strongly pairing to protons occur at higher energy: $E_c \sim E_{\bar{p}} + \alpha_\gamma^{-1} \times (m_{\mathrm{p}} + m_\mu)$ and $E_c \sim E_{\bar{p}} + \alpha_\gamma^{-1} \times (m_{\mathrm{p}} + m_\tau)$ respectively.

In order to compute the critical energy values, we must insert in the expressions not only the current values of the masses of the par-

[29] Keep in mind that, despite their representation in figures 4.6–4.10, these processes are not to be interpreted as depicting Feynman diagrams within a field theory context.

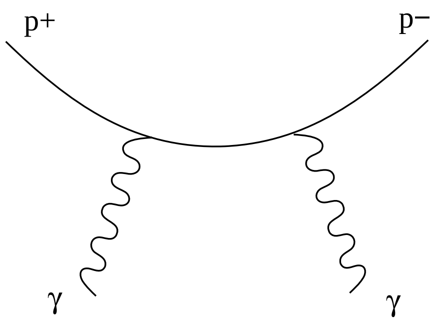


Figure 4.6: Representation of a direct $p\bar{p} \to \gamma\gamma$ scattering channel (the single channels $(u\bar{u} \to 2\gamma, u\bar{u} \to 2\gamma, d\bar{d} \to 2\gamma)$ are here collectively indicated by just one proton line).

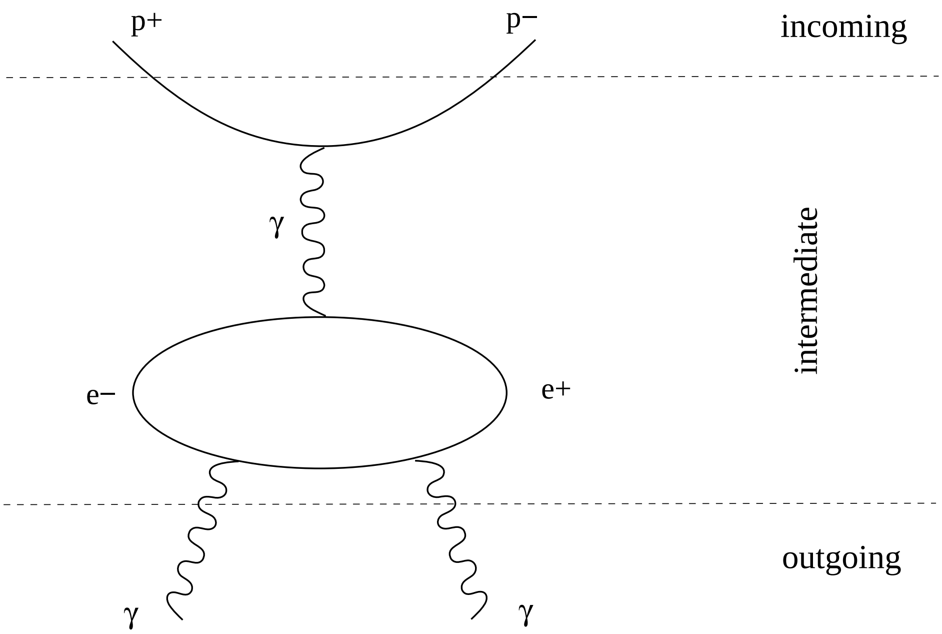


Figure 4.7: Representation of a $p\bar{p} \to \gamma\gamma$ scattering channel via intermediate e^+e^- pair creation.

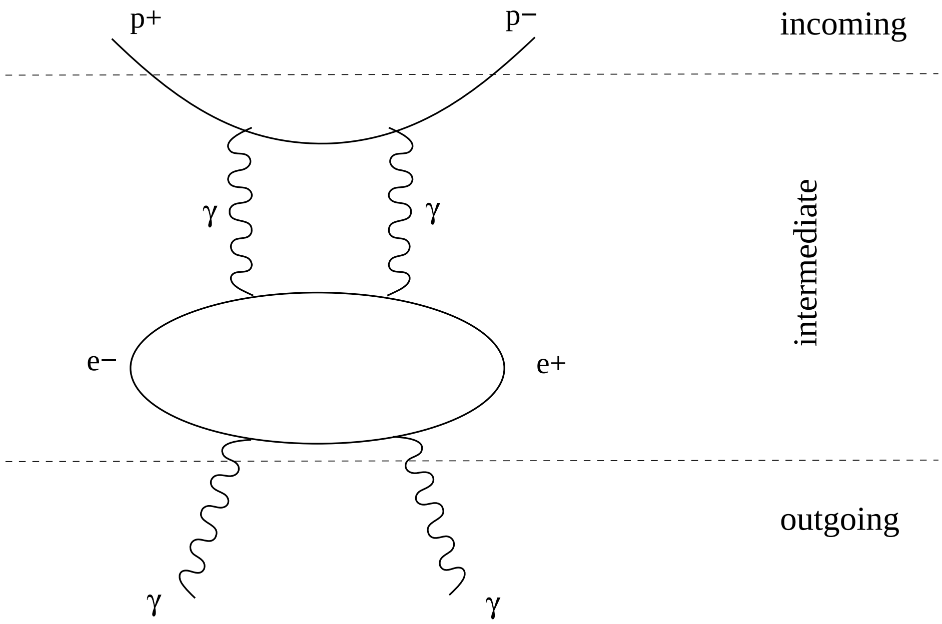


Figure 4.8: Representation of a $p\bar{p} \to \gamma\gamma$ scattering channel via intermediate e^+e^- pair creation.

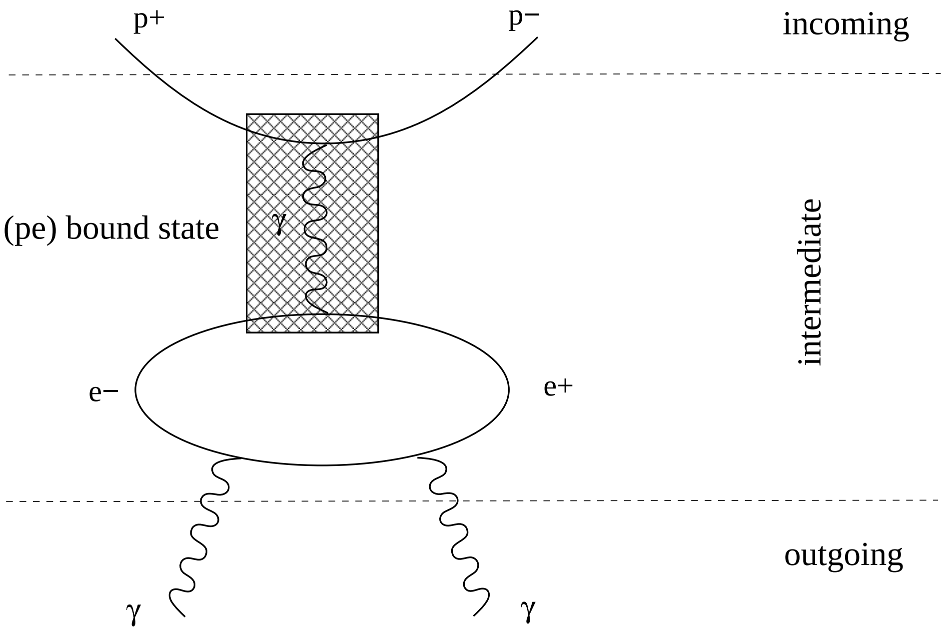


Figure 4.9: Representation of a $p\bar{p} \to \gamma\gamma$ scattering channel via intermediate e^+e^- pair creation at the (pe) bound-state critical energy.

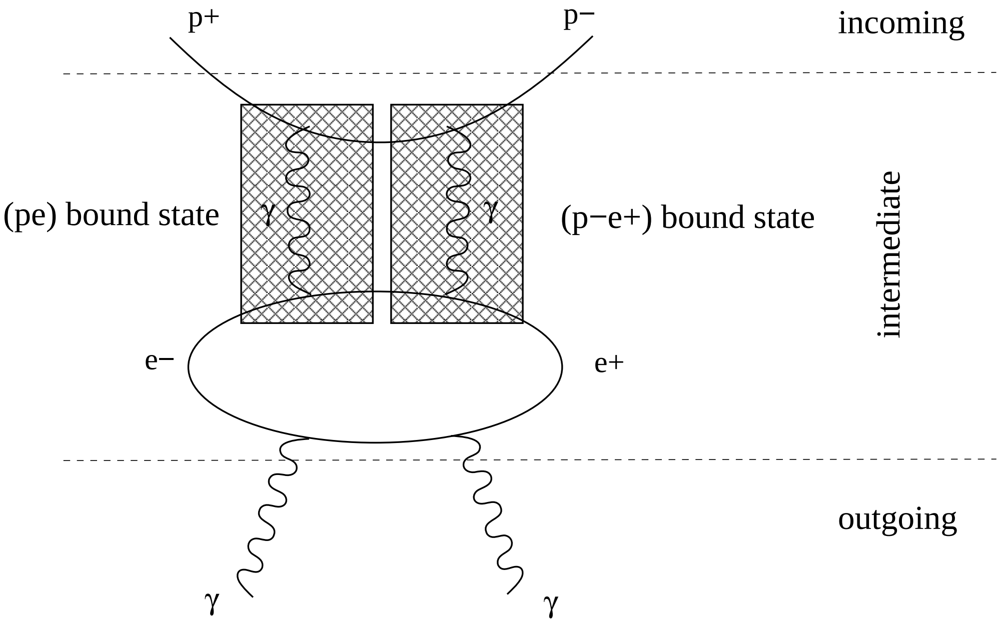


Figure 4.10: Representation of a $p\bar{p} \to \gamma\gamma$ scattering channel via intermediate e^+e^- pair creation at the $(pe)(p^-e^+)$ double bound-state critical energy.

ticles, but also an appropriate value for the electromagnetic coupling. In order to find it we proceed as follows. In all the cases in which a lepton-pair is produced, the process occurs at the level of free particles, namely, it involves just the electromagnetic part of the interaction. For an estimate of the energy at which to run the electromagnetic coupling, we consider therefore the typical energy of the free particles involved, the lepton and the free quarks. In the case of electrons pair, the typical energy of the process is therefore the MeV scale. The electromagnetic coupling is run to this scale according to the behaviour discussed in section 4.3.3, i.e. logarithmically up to the Planck scale, where it is 1. The scale of the heaviest quark is about one order of magnitude higher than the electroweak scale, and 21 orders of magnitude lower than the Planck scale ($\sim 10^{19}\,\mathrm{GeV}$). The value of the inverse coupling at that scale is therefore around $\frac{21}{22} \times 137$ [30]. For this value of the coupling we obtain a critical energy at $\sim 0.939\,\mathrm{GeV} \times 137 \times \mathcal{O}(21/22) + 0.939\,\mathrm{GeV} \approx 124$-$126\,\mathrm{GeV}$. This is only an approximate estimate, the uncertainty depending on our lack of precision in the choice of the energy scale for the computation of the renormalization of the coupling: should it be an average scale between that of the *up* and *down* quarks, or the sum of the quark masses plus the electron mass? The second option, which corresponds to choosing as energy scale the total energy of the involved bare particles, intuitively a reasonable choice, is the one that gives as critical energy 125 GeV. The production of a $\mu\bar{\mu}$ pair occurs at slightly higher energy. Inserting the value of the muon mass, and using for the evaluation of the electromagnetic coupling the 100 MeV scale, we find as energy threshold $E \sim 1.040\,\mathrm{MeV} \times 137 \times \mathcal{O}(20/22) + 1.040 \approx 130$-$131\,\mathrm{GeV}$. The further leptonic peak, corresponding to a $(p\tau)$ state, occurs at much higher energy ($m_\tau \sim 1.777\,\mathrm{GeV}$, implying $\sim 324\,\mathrm{GeV}$ as critical energy). Enhancements of the cross section at higher energies are produced when both the lepton-hadron and the anti-lepton–anti-

[30] We recall that in this theoretical framework the values of masses and couplings are not freely adjustable parameters, but are computed as functions of the only free parameter of this scenario, the age of the universe. Comparison with just one experimental quantity is enough to fix its present value, and to consequently derive the value of all the remaining physical quantities.

hadron pairs are at a virtual strong coupling. These peaks are to be found at about twice the energy of the single-pair peak.

Besides binding quarks with leptons, there is also the possibility of forming intermediate states made of pairs of quarks electromagnetically strongly coupled with other quarks. These can be a subset of the quarks and anti-quarks from the colliding proton-antiproton pair, or pairs formed from quarks of the incoming proton (and/or anti-proton) and virtual quarks instead of virtual leptons. In this case one forms mesonic-like states. The lightest resonances are to be expected from the creation of pion-like bound states, obtained producing intermediate $u\bar{u}$ and $d\bar{d}$ pairs, in which each virtual quark couples electrically to a corresponding quark with opposite charge in the incoming proton or antiproton. Also these states mainly decay into pairs of photons. The evaluation of the energy thresholds is however in this case affected by the fact that now incident and virtual quarks can interact among themselves also through the strong force.

The energy scale of the process, the energy at which the inverse of the electromagnetic coupling must be run, is arguably no more that of the bare quarks. In order to find out what is the right energy scale at which to evaluate the effective electromagnetic coupling to be used in our computations, we must consider that in this theoretical scenario physical parameters are average quantities obtained from a superposition of geometries. In this case, we can figure out that we have a superposition of configurations in which the involved quarks interact partly in triplets to form protons, partly in pairs to form pions. For a rough evaluation of the effective value of the electromagnetic coupling we choose therefore an intermediate scale between the one of the proton and the one of the intermediate meson. Owing to the multiplicative structure of the phase space, we decide for a geometric mean, $\langle E \rangle \approx \sqrt{E_p \times E_\pi}$. This choice should lead us not too far away from the correct value.

We consider now the electric coupling of quarks and anti-quarks. In this case, at the critical energy we gain even powers of the coupling g: 2 or 4, i.e. one or two α_γ^{-1} factors for each quark pair. For instance,

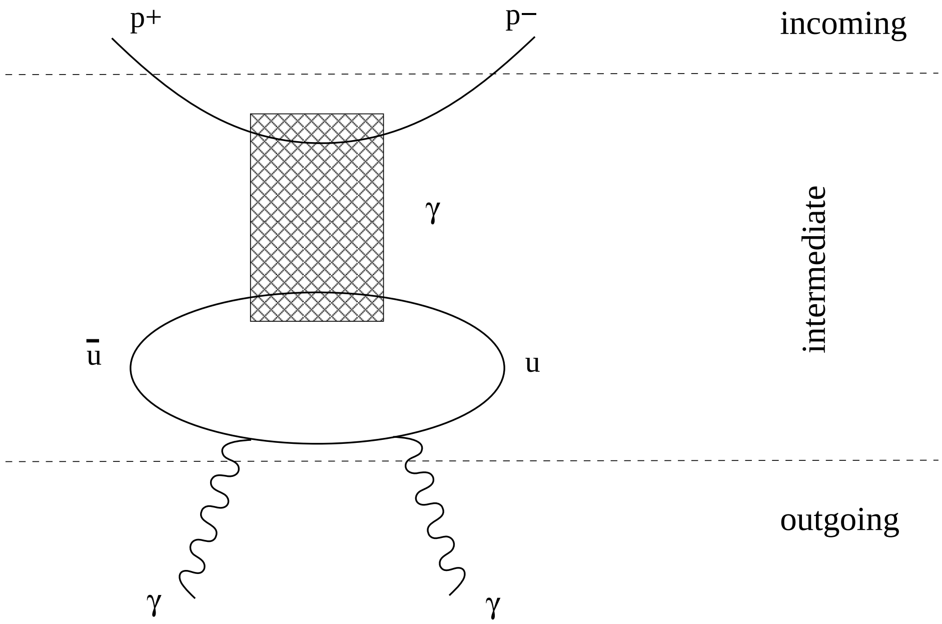


Figure 4.11: Representation of a $p\bar{p} \to \gamma\gamma$ scattering channel via intermediate $u\bar{u}$ pair creation at the pion-like critical energy.

in the case of a $u\bar{u}$ quark pair the analogous of relation 4.5.1 is now a pair of equations:

$$\beta_{u_1} + \beta_{u_2} + \beta_d = -\beta_{\bar{u}} \tag{4.5.5}$$
$$\beta_{\bar{u}_1} + \beta_{\bar{u}_2} + \beta_{\bar{d}} = -\beta_u \,, \tag{4.5.6}$$

where the two gauge parameters on the r.h.s. are not independent. These degrees of freedom are therefore reduced or increased always in pairs. The lowest critical energy is obtained with just one pairing, a configuration that can only occur through a creation of a quark pair, of which only one quark couples with an incident hadron, while the other remains uncoupled. The process is illustrated in figure 4.11. The energy at which this is expected to occur is obtained as:

$$\begin{aligned} &0.942[= m_{\mathrm{p}} + m_{u,d}]\,\mathrm{GeV} \times 137 \\ &\quad \times \mathcal{O}\left(\left[(\log_{10}[(\sqrt{m_{\mathrm{p}}/m_{\pi}}) = 2.6] \;=\; 0.42) + 19 \right] / \, [22] \right) \; [= 120.9] \\ &\qquad\qquad\qquad\qquad\qquad\qquad + \; 0.942\,\mathrm{GeV} \quad \approx \; 114.8\,\mathrm{GeV}\,. \end{aligned}$$

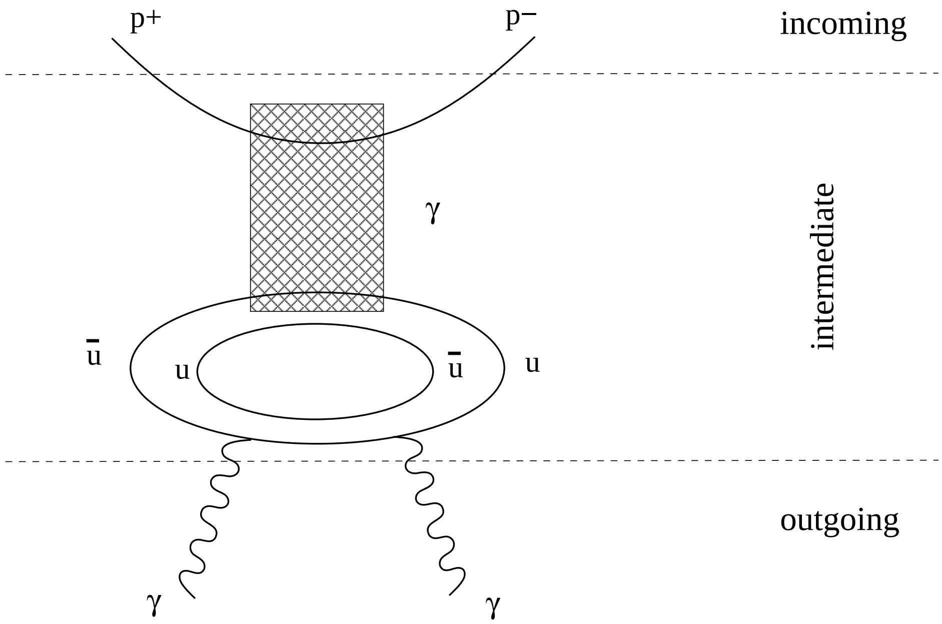


Figure 4.12: Representation of a $p\bar{p} \to \gamma\gamma$ scattering channel via intermediate $\pi\bar{\pi}$ pair creation at the pion-like critical energy.

If in the calculation of the average mass scale the proton mass weights more than what assumed in this computation, one obtains a lower critical energy. The uncertainty in this computation due to the approximation implicit in the choice of the energy scale for the renormalization of the electromagnetic coupling is of the order of 2-3 %, allowing a range of critical energies between some $\sim$ 110-111 GeV to some $\sim$ 115-116 GeV. Notice that we don't need to think that a pair of full pion states is produced. This is an alternative channel, which is illustrated in figure 4.12. In this second case, an analogous computation, taking as starting point the mass of the proton plus the mass of the pion ($\sim$ 129 MeV), gives an enhancement of the cross section at around 130 GeV. A peak at a slightly higher energy is obtained when the quark pair is of the type $s\bar{s}$ (K-like state). In this case, inserting the strange quark mass ($\sim$ 100 MeV) and evaluating the coupling as before, but at an intermediate energy scale between the proton and the K-meson, we obtain $1.043\,\text{GeV} \times 137 \times \mathcal{O}(19.12/22) + 1.043\,\text{GeV} \approx 125\,\text{GeV}$. This concurs to increase the strength of the enhancement around 125 GeV.

Considering instead as evaluation scale for the electromagnetic coupling an analogous average scale, but this time with the average taken between the proton mass and the mass of the bare s-quark, one obtains a peak close to 130 GeV. If one further takes into account the possibility of producing not only the $s\bar{s}$ quark pair, but a whole K-meson pair, one gets a peak at an energy about 50% higher than these energy scales. Along the same line, one can compute the critical energies for the enhancements occurring at a higher scale, produced by $c\bar{c}$, $b\bar{b}$ and $t\bar{t}$. They are expected to show up respectively at $\approx 266\,\text{GeV}\,[p-c]$, $\approx 594\,\text{GeV}\,[p-b]$ and $\approx 1.8 \times 10^4\,\text{GeV}\,[p-c]$ [31]. In the case of strong coupling among quarks of the colliding hadrons, the quark on the r.h.s. of 4.5.6 forcedly coincides with one of those on the l.h.s. of 4.5.6. The two equations are therefore always coupled and the situation is equivalent to a double pairing, leading to an α_γ^{-2} volume enhancement factor, at a much higher critical energy [32].

No enhancement of this type is expected to occur when the intermediate pair produced in the scattering consists of charged pions. In the case of $p\bar{p}$ scattering the quarks of the intermediate pions re-combine with those of the proton and anti-proton to give rise once again to pions, produced through a rearrangement of the degrees of freedom. In the case of lepton-antilepton scattering, there is no possibility of forming pairs with the quarks of the pions which, at the strong coupling, can lead to a reduction of the electromagnetic gauge symmetry. This is due to the fact that pions, either neutral or charged, are made of quark-antiquark pairs (e.g. $\pi^+ \leftrightarrow u\bar{d}$). The $SU(2)$ symmetry relating *up* and *down* in this scenario is broken by the introduction of

[31] In detail: $m_c = 1.29\,\text{GeV} \Rightarrow$ proton-charm $\to (0.938 + 1.29) \times 137 \times \mathcal{O}[19/22] + 2.228 \sim 266\,\text{GeV}$; $m_b = 4.18\,\text{GeV} \Rightarrow$ proton-bottom $\to (0.938 + 4.18) \times 137 \times \mathcal{O}[18.5/22] + 5.118 \sim 594\,\text{GeV}$; $m_t = 173.3\,\text{GeV} \Rightarrow$ proton-top $\to (0.938 + 173.3) \times 137 \times \mathcal{O}(18/22) + 174.238 \sim 19.705 \approx 1.8 \times 10^4\,\text{GeV}$.

[32] For the purpose of determining the scale of the critical energy it is not so relevant to decide if one has to add to the computation the mass of the free virtual quark or the one of the meson (it is a matter of 100 MeV's order till 1-2 GeV as compared to the 100 and more GeV). It matters if it has to be included in the multiplicative rescaling through α_γ^{-1} factors. If we do this in the case of pions we obtain $1.070 \times 137 \times (19/22) + 1.070 \approx 128$, a contribution which is going to increase, and widen out, the peak around 130 GeV.

masses. Since these run as a power of the inverse of the age of the universe, $m \sim 1/\mathcal{T}^p$ for appropriate exponents p, at the present conditions of the universe, i.e. at large age/volume ($\mathcal{T} \gg 1$), its breaking can be considered a kind of "soft breaking". On the contrary, the gauge parameters of the electromagnetic gauge group, and in particular relations like 4.5.1 involving the breaking into quarks and leptons, are scale-insensitive. As a consequence, for the gauge parameters of the electromagnetic group the separation between families of particles and, inside each family, between $SU(2)$ doublets, are second-order effects: in first approximation the *up* and *down* of each doublet are to be considered the same kind of particle, simply with a different charge. Also a $\bar{d}$ quark is like an anti-u quark, simply with a different normalization of the charge. Since all mesons are of the type $q_i \bar{q}_j$, where i and j run over the families of quarks and the two values indicating the upper and down members of an $SU(2)$ pair, for the sake of the present analysis they can all be considered of the type $q\bar{q}$, i.e. consisting of a quark and its anti-quark. For all of them, the charge neutrality condition analogous to 4.5.1 is of the type $\frac{1}{3}\beta + \frac{2}{3}\beta = \beta$, with *just one parameter*, as is the case of a lepton-antilepton pair. This implies that already at the weak coupling the set of particles of the pair do transform under $U(1)$ *all together*, as if they were one single particle. The analogous of relation 4.5.1 does not involve in this case free parameters, and there is no volume group factor to be lost at the strong coupling. For what matters the number of gauge parameters, there is therefore no difference between weak and strong coupling, and we expect no enhancements of the cross section due to meson-lepton bound states to occur.

To summarize, there are several configurations concurring to enhance the $\gamma\gamma$ decay channels, spread out in an energy interval going from $\sim 111\,\text{GeV}$ to $\sim 130\,\text{GeV}$, with some peaks around 111-115 GeV, 125 GeV, and 130 GeV. Further enhancements are to be found at higher energies. Since in these processes we don't deal with diverging quantities, each channel contributing to this kind of resonance is expected to enhance the decay amplitude by a relatively small amount.

Its effect may therefore be difficult to detect and identify out of the ground decay channels and the statistical noise fluctuations, unless there are several peaks close enough to each other, so that their widths can overlap. Around 125 GeV there is indeed a whole bunch of configurations with peaks potentially overlapping due to their statistical width. They must be compared with the resonance found in the $p\bar{p}$ scattering at LHC [60, 78, 79], which has an analogous signature. This is the energy at which this effect is expected to manifest itself in the strongest way. Besides this line, at a lower level of strength we expect to find the line around 130 GeV, which is also the result of a collection of contributions. Although apparently not detected in the Large Hadron Collider, this threshold could be the line observed by astronomers [80, 81, 82, 83], that in our framework is therefore not interpreted as an evidence of dark matter. Astrophysical observations give indications also for a line around 111 GeV [84], which could be compatible, within the approximations implied in our computations, with the enhancement at 111-115 GeV we have found as first energy threshold.

5 Cosmology

5.1 The geometry of the universe

As discussed in chapter 3, the absence in our theoretical framework of symmetry under space-time translations implies a different normalization of string amplitudes, which must be now normalized in such a way that densities scale like the inverse of the Jacobian of the transformation between string world-sheet and target space coordinates. An amplitude which in the light-cone gauge is of order one, like the vacuum energy in the non-supersymmetric orbifolds considered in the previous section, in which supersymmetry is broken at the unit scale (identified with the Planck scale), gives therefore an energy density which scales as:

$$\rho(E) \; \sim \; \frac{1}{\mathcal{T}^2} \,. \tag{5.1.1}$$

In order to get the value of a global quantity, like the entropy, we must instead multiply the string amplitude by the Jacobian factor, obtaining the scaling:

$$S \; \propto \; \mathcal{T}^2 \,. \tag{5.1.2}$$

The total energy at a certain time $\mathcal{T}$ of the history of the universe, given by the integral of the energy density over the space volume of the universe at time $\mathcal{T}$, scales then as:

$$\mathrm{E}(\mathcal{T}) \; \sim \; \int_{\mathcal{T}} d^3 \frac{1}{\mathcal{T}^2} \;\; \approx \; \mathcal{T} \,. \tag{5.1.3}$$

In the string representation we recover therefore the values we computed in the ground description of this scenario.

5.1.1 The solution of the FRW equations

The density 5.1.1 collects both the pure geometric, i.e. cosmological, and the matter/radiation contribution to the energy density. These terms are separately of the same order. The reason is that the set of most singular string vacua inherits what remains of the symmetry under exchange of three sectors of the theory at the $\mathcal{N}_4 = 2$ level, the $S - T - U$ symmetry of the orbifold construction, which can be seen to exchange the roles of gravity, matter and radiation by exchanging the sectors giving rise to the corresponding fields. In the further steps of symmetry breaking this symmetry is broken by terms of order $\mathcal{O}(1/\mathcal{T}^p)$ in the string partition function. The energy densities get therefore distinguished by higher order terms: $\rho \sim \frac{1}{\mathcal{T}^2} \longrightarrow \frac{1}{\mathcal{T}^2}\left(1 + \mathcal{O}(1/\mathcal{T}^p)\right)$.

Let us now investigate the geometry of the expansion of the universe. As the universe evolves, the energy density and the curvature of space-time decrease toward a flat limit, and the dominant configuration tends to a "classical" description. At large $\mathcal{T}$ it is therefore reasonable to suppose that this configuration admits a description in terms of Robertson-Walker metric, i.e. a classical metric of the type:

$$ds^2 \, = \, dt^2 \, - \, R^2(t) \left[\frac{dr^2}{1 - kr^2} \, + \, r^2 (\, d\theta^2 \, + \, \sin^2 \theta \, d\phi^2) \right] \, , \qquad (5.1.4)$$

where, in our case, $t \equiv \mathcal{T}$, and $r \, \leq \, 1$. The metric should correspond to a closed universe, $k = 1$. Under the assumption of perfect fluid for the energy-momentum tensor, the Einstein's equations lead to:

$$\left(\frac{\dot{R}}{R} \right)^2 \, = \, - \, \frac{k}{R^2} \, + \, \left\{ \frac{8\pi \, G_{\rm N} \, \rho}{3} \, + \, \frac{\Lambda}{3} \right\} \, , \qquad (5.1.5)$$

where we have collected within brackets the contribution of the stress-energy tensor and of the cosmological term. Inserting the "Ansatz" $R \, = \, \mathcal{T}$ we obtain:

$$\left(\frac{\dot{R}}{R} \right)^2 \, = \, - \, \frac{(k = 1)}{R^2} \, + \, \left\{ \frac{\sim 2}{R^2} \right\} \, \sim \, \frac{1}{R^2} \, , \qquad (5.1.6)$$

that we can write as:

$$\left(\frac{\dot{R}}{R} \right)^2 = \frac{\kappa^2}{R^2} \,, \tag{5.1.7}$$

for some coefficient κ. The equation is solved by $R = \kappa\, t$, consistently with our Ansatz. This confirms that the dominant configuration corresponds to a spherical Robertson-Walker metric, describing a universe bounded by a horizon expanding at a fixed ratio to the speed of light.

The comparison of our results with astronomical data contains however a possible weak point. Experimental data are given as a result of a process of interpretation of certain measurements, for instance through a series of interpolations of parameters. All this is consistently done within a well defined theoretical framework. Usually, one takes a "conservative" attitude and lets the interpolations run in a class of models. However, this is always done within a finite class of models. In principle, we are not allowed to compare theoretical predictions with numbers obtained through the elaboration of measurements in a different theoretical framework: in general, this doesn't make any sense. However, in the present case this comparison is not meaningless, and this not on the base of theoretical grounds: the reason is that, for what concerns the time dependence of cosmic parameters and energy densities, the solution we are proposing does not behave, at present time, much differently from the "classical" cosmological models usually considered in the theoretical extrapolations from the experimental measurements. The rate of variation of energy density is in fact: $\dot{\rho} \sim \partial(1/R^2)/\partial\mathcal{T} = 1/\mathcal{T}^3 = 1/R^3$. The values of the three kinds of densities can therefore be approximated by a constant within a wide time interval. For instance, as long as the accuracy of measurements does not go beyond the order of magnitude, these densities can be assumed to be constant within a range of several billions of years. For the purpose of testing the statements and conclusions of the present analysis, the use of the known experimental data about the cosmological constant, derived within the framework of a Robertson-Walker universe with constant densities, is therefore justified.

A universe evolving according to eq. 5.1.6 is *not accelerated*: $\dot{R} = 1$

and $\ddot{R} = 0$. Owing to the existence of an effective Robertson-Walker description, the red-shift can be computed as usual. We have:

$$1 + z = \frac{\nu_1}{\nu_2} = \frac{R_2}{R_1} = \frac{\mathcal{T}_2}{\mathcal{T}_1}, \tag{5.1.8}$$

where ν_1 is the frequency of the emitted light, ν_2 the frequency which is observed, and R_1, R_2 are respectively the scale factor for the emitter and the observer. $R = \mathcal{T}$ is precisely the statement that the expansion is not accelerated. Expression 5.1.8 however accounts for just the "bare" red-shift, namely the part due to the expansion of the universe: it does not account for the corrections coming from the time dependence of masses and couplings, that we will discuss in section 5.1.2. Usually, this effect is not taken into account, because in the standard scenarios masses are assumed to be constant. In our scenario they depend instead on the age of the universe. A change in the values of masses and couplings reflects in a change of the atomic energy levels, and therefore in a change of the emitted frequencies. We will see that, once the observed frequencies in expression 5.1.8 are corrected to include also the change in the scale of energies, the scaling of the emitted to observed frequency ratio is not anymore proportional to the ratio of the corresponding ages of the universe. Since the conclusions about the rate of expansion are precisely derived by comparing red-shift data of objects located at a certain space-time distance from each other, this explains why the expansion *appears to be accelerated.*

5.1.2 The apparent acceleration of the universe

We are now in a position to come back to the issue of the apparent acceleration of the universe. In our framework, atomic energy levels depend on the age of the universe. More precisely, each energy level scales in principle as a different function of time. Their ratios, and therefore the ratios of emitted frequencies, are not constant over time. We can separate the time dependence into an overall average effect, a time-dependent set of "central values" of ratios, expressed as a unique function of time, common to all atomic spectra, and time-dependent departures from this central value, that take into account the independent scaling of each energy level. The overall average effect is what results in an effective red-shift, whereas the second term, the individual departures, are responsible for what in the literature, depending on the model and approach, are referred to as "time variation of α", or "time variation of the mass", or "time-dependent relativistic effects". We will come back to this issue in section 5.4.0.1. Here we consider the universal term. The main contribution to the time-dependence of the atomic spectra comes from the dependence of the Bohr radius on the quantity $m\alpha^2$, where m is the electron's reduced mass, and α the electromagnetic coupling. In order to give a rough estimate of the red-shift effect which is produced, we can approximate the time dependence of any mass with the one of the stable matter scales:

$$m \sim \mathcal{T}^{-3/10}. \tag{5.1.9}$$

The time-dependence of α can be obtained from the expressions given in section 4.2.1.2, and turns out to be $\alpha \sim \mathcal{T}^{-\frac{47}{28\times 45}}$, negligible as compared to the time dependence of the mass. From this, we derive that the above behaviour induces an apparent shift in the frequencies of the light emitted at different distances from the observer, i.e. at different ages of the universe, due to the different scale of the atomic energy levels, of the order:

$$\frac{\tilde{\nu}_1}{\tilde{\nu}_2} = \left(\frac{\mathcal{T}_2}{\mathcal{T}_1} \right)^{\frac{3}{10}}. \tag{5.1.10}$$

Once "subtracted" from the bare red-shift 5.1.8, this gives an apparent, effective red-shift $z_{\rm app.}$:

$$1 + z_{\rm app.} = \left(\frac{\nu_1}{\nu_2} \right)_{\rm observed} = \left(\frac{\mathcal{T}_2}{\mathcal{T}_1} \right)^{\frac{7}{10}}, \tag{5.1.11}$$

as if the universe were expanding with rate $\tilde{R} \sim \mathcal{T}^{7/10}$, normally expected for a matter dominated era.

At the base of what is considered an experimental evidence of the accelerated expansion of the universe is the observed acceleration in the time variation of the red-shift effect. In the classical approach, the expansion occurs at the level of the overall scale factor of the space part of the Robertson-Walker metric:

$$ds^2 \;=\; dt^2 \,-\, R^2(t) \left[d\vec{x}^2 \right] \,. \tag{5.1.12}$$

One must however underline that this is the metric ruling the cosmological scale of the universe. If the rescaling expressed in 5.1.12 was instead valid at any scale, it would imply a change in the overall scale of physics. In practice, just an unobservable scale redefinition [1]. In our approach, the red-shift effect receives a different explanation, being given in terms of accelerated variation of ratios of mass and energy scales, and therefore of observed emitted frequencies, without recourse to an accelerated expansion of the metric at a cosmological scale. There is therefore no need for a conceptual separation between a local, and an effective, large-scale description of physics.

[1] The scale factor $R(t)$ precisely defines the speed of light (obtained from the condition $ds^2 = 0$, which implies $dx/dt = 1/R$). Saying that there is an expansion of the overall scale of the metric is equivalent to saying that there is an expansion of the scale according to which space lengths are measured in terms of time length. In other words, saying that there is such an expansion means that there is an expansion (more precisely a contraction) of the speed of light. Suppose we want to compare wavelengths between present time and a time at which the scale was 1/2 of the present one. From a physical point of view, what we observe is radiation produced by atomic transitions, and we compare wavelengths keeping fixed the period of the light wave. Since in the past time lengths were contracted by 1/2 with respect to today, during each period of the wave light was traveling twice as much as today. Therefore, the same atomic transition generated a photon with twice the wavelength as today. However, if the space scale was contracted, also energies were different. Energies scale in fact as inverse of lengths (consider for instance the electric potential, $V = e^2/R$). In our specific example, this means that energies were doubled, and, according to $E = h\nu$, also frequencies were doubled, or equivalently periods were halved. The same atomic transition produced therefore photons with twice the frequency, or half the period, as compared to today. This fact, combined with the fact that the speed was doubled, implies that, for the same physical phenomenon, the effective wavelength was the same as today. Any such an overall scale of the metric would be physically unobservable.

5.2 The CMB radiation

In the usual cosmological interpretation, the cosmic background radiation, which has the typical spectrum of a black body radiation with a temperature of about 2.8 K [85, 86], is interpreted as being the remnant of very early processes in the universe. It would consist of photons cooled down during the expansion of the universe. At the origin they should have possessed an energy corresponding to a microwave length, as expected from energy exchange due to Compton scattering through the plasma at the origin of the universe. The low temperature would then be the effect of the cooling down of the universe due to its expansion.

In our theoretical framework it is not necessary to advocate the primordial history of the universe in order to account for the existence of a low-temperature radiation. Being a background radiation, it must not evidently come from clearly identified sources such as electronic transitions in the elements composing stars etc. Indeed, the fact that the superposition of configurations 3.1.4 leads to a spectrum that we can interpret in terms of the usual elementary particles and fields does not mean that the physics of the universe is completely accounted in terms of these degrees of freedom and their interactions. Like the masses of the elementary particles, also the photon energies are the result of an averaging procedure over all the configurations. As such, they do not necessarily correspond to energy levels of ordinary elementary particles. In section 4.1.1.4 we have seen that all massive states are built over a background corresponding to the (s, s, s) configuration. We may think of fluctuations around this background. For instance, electron-positron pairs that are temporarily popped out, and disappear into a pair of photons. In order to estimate the mean energy of such a radiation, we may think of the background as a kind of thermal bath, constituted by "particles" of mass:

$$\langle m \rangle \ \sim \ \frac{1}{\sqrt{\mathcal{T}}} \, . \tag{5.2.1}$$

The normalization is twice as much as M_0 as given in 4.1.12, because here we are looking for neutral states, therefore possible particle-antiparticle pairs. We can obtain the energy of a radiated photon by considering once again relation 4.3.44, where this time instead of the square of the W mass we have the square of the energy of the photon E_γ, and on the r.h.s. we have the electromagnetic coupling α_γ and the mass $\langle m \rangle$:

$$\langle E \rangle_\gamma^2 \;\sim\; \alpha_\gamma \langle m \rangle \langle m \rangle \,, \tag{5.2.2}$$

from which we obtain:

$$\langle E_\gamma \rangle \;\sim\; \sqrt{\alpha_\gamma}\, \frac{1}{\sqrt{\mathcal{T}}} \,. \tag{5.2.3}$$

This expression can be interpreted in the following way: the mean energy of the radiated photons is not exactly the mean ground energy, because the average is weighted by the fraction of phase space volume which is effectively involved in the electromagnetic interaction with the background. This fraction is precisely set by the value of the electromagnetic coupling. The scale at which α_γ is evaluated is not necessarily M_0: if the radiation is produced by electron-positron interactions, the appropriate scale could be the rest energy of the electron-positron system, or lie something above it. Just to be concrete, if we insert in 5.2.3 the value of α_γ at the electron's scale, as derived through a logarithmic running from M_0 as in section 4.3.3, $\alpha_\gamma^{-1}|_{m_e} \sim 131.4$, and the value A.1 for the present age of the universe, after converting energy into temperature through the Boltzmann constant we obtain:

$$T_\gamma \;\equiv\; k^{-1} < p_\gamma > \, = \, k^{-1} E_\gamma^0 \;\sim\; 2.70\,\mathrm{K} \,. \tag{5.2.4}$$

If we instead run the coupling to an average $\langle m_e m_u m_d \rangle$ scale, we have $\alpha_\gamma^{-1}|_{m_e} \sim 132.8$, and we obtain:

$$T_\gamma \;\sim\; 2.73\,\mathrm{K} \,. \tag{5.2.5}$$

The Gaussian tail of the resonance, leading to a black-body distribution of frequencies, is in this context the consequence of the superposition 2.1.16, for which the entropy sum, once restricted to the phenomenon under consideration, and thermodynamically interpreted as in section 3.5, namely through $S \sim E/T$, becomes a typical Gaussian distribution.

5.3 The fate of dark matter and the Chandra observations

A discrepancy between our framework and the common expectations is the absence in our scenario of dark matter. According to our analysis, the universe consists only of the already known and detected particles. Of course, there can be regions of the space in which a high concentration of neutrinos, which for us are massive, increases the curvature without being electromagnetically detected. But this is not going to change dramatically the scenario: there is no hidden matter acting as an extra source able to increase the gravitational force by around a factor ten over what is produced by visible matter, as it seems to be required in order to explain a gravitational attraction among galaxies much higher than expected on the base of the estimated mass of the visible stars. The problem arises in several contexts: Big Bang nucleosynthesis, rotational speed of galaxies, gravitational lensing. All these points would require a detailed investigation, beyond the scope of this work. We will also not attempt to rediscuss a huge literature, and limit ourselves here to mention some hypotheses. The first remark is that the discrepancies between theoretical expectations and the observed effects, which are found in so different issues as primordial universe, nucleosynthesis and galaxy phenomenology, don't need necessarily to be explained all in the same way.

Let's consider the problems related to the motion of external stars in spiral galaxies, where for the first time the issue of dark matter has been addressed, and the "anomalous" gravitational lensing, with reference to the effect observed in the 1E0657-558 cluster [87]. It is since 1933 (Fritz Zwicky) that, by looking at the amount of red-shift in the light emitted by the stars in the wings of a spiral galaxy, it has been noticed how, differently from what expected, the rotation speed does not decrease with the inverse of the square root of the radius: it is a constant [88, 89]. Presence of invisible matter has been advocated, in order to fill the gap between the mass of the observed matter and the amount necessary to increase the gravitational force. Indeed, the expectation that the rotation speed of stars in the external legs should decrease is based on the assumption that almost the entire mass of the galaxy is concentrated in the bulge at the center of the spiral. Any star on the wings would therefore feel the typical gravitational field due to a fixed, central mass.

In the framework of our scenario, masses have been in the past higher than what they are now. Moreover, owing to the fact that, as we discuss in chapter 2, the universe "closes up", in such a way that the horizon we observe corresponds to a "point", the space separation between objects located at a certain cosmic distance from us appears to be larger than what actually is. All this could mean that the mass of the center of a galaxy, as compared to the wings, has been systematically overestimated. It would be interesting to see, by carrying out a detailed re-examination of the astronomical observations, whether the behaviour of the center of a galaxy still requires to advocate the presence of a heavy black hole, in order to explain a gravitational force higher than what expected on the base of the estimated mass of the visible stars. In any case, it is possible that, once the downscaling of length and upscaling of masses has been appropriately taken into account, a better approximation of a spiral galaxy is the one sketched in figure 5.1. In part A of the picture the galaxy is (very roughly) represented with wide wings, with stars relatively "broadened" on the plane of the galaxy. Part B shows the same figure, simply with much narrower arms. In picture A the broad lines have been shadowed in a way to make evident that the higher star density of the bulge is largely due to the "superposition" of the various arms. Nevertheless, as it is clear from picture B, the problem remains basically "one-dimensional": the wings are one-dimensional lines coming out of the center of the galaxy. Under the hypothesis that all the stars have the same mass, the linear density of a wing is constant, and, once integrated from the center up to a certain radius R, the total mass M_R of the portion of galaxy enclosed within a distance R from the center is roughly proportional to R:

$$\rho = \frac{dM}{dr} \sim \text{const.} \quad \Rightarrow \; M_R \sim \text{const} \times R \,. \tag{5.3.1}$$

In the expression of the gravitational potential, the linear R dependence of the mass cancels against the R appearing in the denominator (the potential remains the one of a Coulomb force). The whole galaxy is the superposition of several pieces of this kind. The gravitational potential energy is therefore a constant times the mass of the star in the wing. Conservation of energy implies therefore that also the velocity of the star does not depend on the radius R. We stress that this is only an approximation: it would be exact if the arms were not those of a spiral but straight legs coming out radially from the center, and under the assumption that all the stars of the bulge correspond to the superposition of the arms.

In the case of the 1E0657-558 cluster, the Chandra observatory has detected a gravitational lensing higher than what expected on the base of the amount of luminous matter. Moreover, the highest effect corresponds to two dark regions close to the cluster, rather than to places where the visible matter is more dense. In the framework of our scenario, a possible explanation could be that what is observed is the effect of a "solitonic" gravitational wave, produced as a consequence of the separation of one sub-cluster from the other one. This could increase the gravitational force by an amount equivalent to the displaced cluster mass, for a length/time comparable to the cluster size, therefore a time much higher than the few hours during which the effect has been measured ($\sim$ 140 hours). It remains that the lensing is around 8-9 times higher than what expected on the base of the amount of visible mass. However, the cluster under consideration is at about 4 billion light years away from us. This is around 1/3 of the age of the universe. This time distance is large enough to make relevant the effects due to a change of the curvature of space-time along the evolution of the universe, as well as the change of masses. Furthermore, as we discussed above, the apparent space separation between objects located at a certain cosmic distance from us must be appropriately downscaled, in order to account for the curving up of space-time into a sphere, with the horizon "identified" with the origin. Putting all this together, we obtain that the effective gravitational force experienced on the 1E0657-558 cluster is (or, better, was) indeed 8-9 times higher than what it appears to us on the base of the expected mass of the objects in the cluster. This is precisely the amount otherwise referred

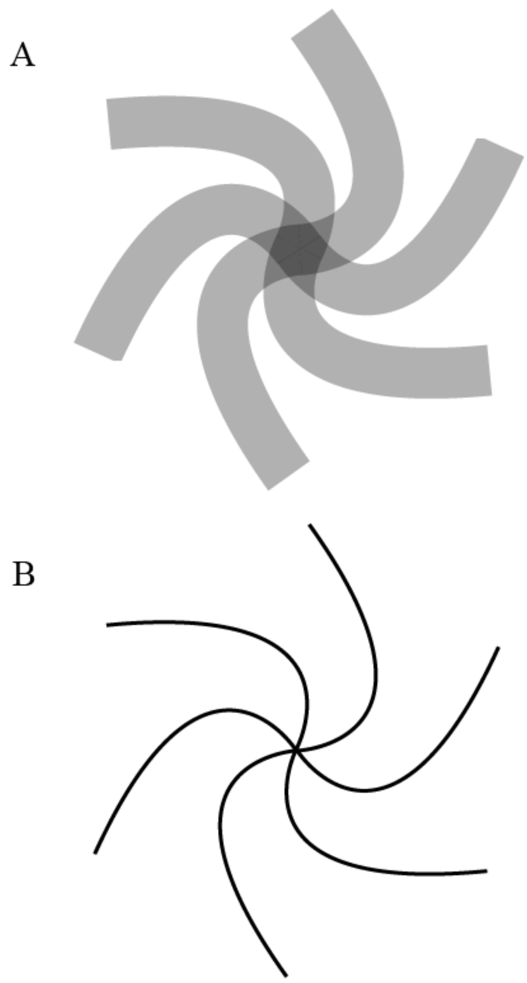


Figure 5.1: Picture A is the rough sketch of a spiral galaxy, in which the arms are broad and shadowed in a way to highlight the increasing mass density due to their superposition at the center. Figure B represents the same object, with the arms narrowed down, in order to highlight the one-dimensional nature of the physical problem, for what concerns the mass density.

to dark matter.

5.4 Cosmological constraints

Cosmology addresses two kinds of problems for what concerns the "running back" of a theory, or an "early time" model. Namely, i) the possible non-constancy of what are commonly called "constants", and ii) the agreement with the expected origin/evolution of the early universe (baryogenesis, nucleosynthesis etc...). In our framework, these issues are put in a light quite different from the usual perspective: there are in fact indeed no constants; therefore, a variation of couplings, masses, cosmological parameters, and, as a consequence, energy spectra, is naturally implemented. However, there is a peculiarity: all these parameters scale as appropriate powers of the age of the universe. As a consequence, a "number" close to one at present day has a very mild time dependence:

$$\mathcal{O}(1) \approx \mathcal{T}^{\epsilon} \;\Rightarrow\; |\epsilon| \ll 1 \,, \tag{5.4.1}$$

and therefore varies quite a little with time. Oklo and nucleosynthesis bounds, being given as ratios of masses and couplings that cancel each other to an almost "adimensional" quantity, are precisely of this kind. In our case they don't provide therefore any dangerous constraint.

For what concerns the non-constancy of "constants", there are not enough data enabling to test our prediction about a time variation of the cosmological constant, whose measurement is still too imprecise. A more stringent test of the variation of parameters comes from the observations on the light emitted by ancient Quasars. In this case, the spectrum shows an "anomalous" red-shifted spectrum. This shift should not be confused with the usual red-shift, of which we have discussed in section 5.1.1. The effect we consider here persists once the "universal" red-shift effect has been subtracted. As an explanation, it is often advocated a possible time variation of the fine-structure constant α.

5.4.0.1 The "time dependence of α"

The issue of the possible time variation of the fine-structure constant arises in the framework of string theory derived effective models for cosmology and elementary particles. Various investigations have considered the possibility of producing some evidence of this variation, or at least a bound on its size. To this regard, astrophysics is certainly a favoured field of research, in that it naturally provides us with data about earlier ages of the universe. A possible signal for such a time variation could be an observed deviation in the absorption spectra of ancient Quasars [90, 91, 92, 93]. This effect consists is a deviation in the energies corresponding to some electron transitions, which remains after subtraction of the background effect of the red-shift, and is obtained with interpolations and fitting of data.

What is observed is a decrease of the relativistic effects in the energies of the electrons cloud, with respect to what expected on the base of present-day parameters (in particular, the fine-structure constant). Indeed, while the atomic spectra are universally proportional to the atomic unit $me^2 \propto m\alpha^2$, the relativistic corrections depend on the coupling α. After subtraction of the "universal" red-shift effects, their variation should then be directly related to a variation of α. In our framework, the explanation comes from considering both the scaling of α and the one of masses at the same time: going backwards in time, α increases, as also the proton and the electron mass do, but *the ratio of α to the mass scales decreases*. Namely, if we measure the variation of α with respect to the electron mass scale (whether the true electron mass or the reduced mass doesn't make a relevant difference [2]) we observe a decrease of the coupling α. Indeed, the main term contributing to the time dependence of atomic spectra is the product $m\alpha^2$, entering the expression of the Bohr radius. If we define a reduced coupling as the coupling measured in units of $m\alpha^2$ we can explicitly see that it

[2] In the hydrogen atom this is given by $\mu = m_e m_\mathrm{p}/(m_e + m_\mathrm{p})$. A discussion about the possibility of referring to a change of this quantity the effect measured in ref. [91] can be found in refs. [94, 95, 96].

decreases if we go backwards in time:

$$\bar{\alpha} \overset{\text{def}}{\equiv} \frac{\alpha}{m\alpha^2} \approx \mathcal{T}^{\frac{1}{3}+\frac{1}{28}} . \tag{5.4.2}$$

The results reported in the literature (see ref. [91]) exclude from the evaluation the effect of the red-shift. Atomic energies have an approximate scaling of the type (see for instance ref. [91]):

$$E_n \approx K_n \, (m\, \alpha^2) + \Gamma_n \, \alpha^2 \, (m\, \alpha^2) , \tag{5.4.3}$$

where K_n and Γ_n are constants and the second term, of order α^2 with respect to the first one, accounts for the relativistic corrections.

In our case all energies scale with time. The red-shift is an average effect, based on a central value of the atomic spectra, out of which depart deviations due the different time-scaling of the various energy levels, which are different functions of mass and coupling. They can be considered of order α^2 as compared to the central value. The quantity suitable for a comparison is therefore:

$$\begin{aligned} \langle |\Delta (E - \langle E \rangle)| \rangle &\sim \mathcal{O}(\alpha^2) \times \partial_t \ln \left(\frac{\alpha^2}{m\alpha^2} \right) \times \Delta t \\ &\sim \mathcal{O}(\alpha^2) \times \frac{3}{10} \times \left(\frac{1}{5} \times 10^{-60} \right) \times \left(4 \times 10^{50} \right) \ \mathrm{yr}^{-1} \\ &\sim \mathcal{O}(10^{-15}) \ \mathrm{yr}^{-1} , \end{aligned} \tag{5.4.4}$$

where the last factor accounts for the conversion of time from Planck units to years (see appendix). This is the relative variation of the relativistic correction subtracted of the universal part (reabsorbed in the red-shift), to be compared with the results of [90], as reported also in [91]:

$$\frac{\langle \dot{\alpha} \rangle}{\alpha} = -2.2 \pm 5.1 \times 10^{-16} \ \mathrm{yr}^{-1} = \mathcal{O}(10^{-15}) \ \mathrm{yr}^{-1} . \tag{5.4.5}$$

5.4.0.2 The Oklo bound

Data from the natural fission reactor, active in Oklo around two billions years ago, are today considered one of the most important sources of constraints on the time variation of the fundamental constants. By comparing the cross section for the neutron capture by Samarium at present time with the one estimated at the time of the reactor's activity, one derives a bound on the possible variation of the fine-structure constant, and on the ratio $G_F m_{\mathrm{p}}^2$, in the corresponding time interval. The interpretation of the experimental measurements and their translation into a bound on the variation of the capture energy resonance is not so straightforward, and depends on several hypotheses. In any case, all these steps are sufficiently under control. More uncertain is the translation of this bound on the energy variation into a bound on the variation of the fine-structure constant and other parameters: this passage requires strong assumptions about what is going to contribute to the atomic energies. This analysis was carried out in ref. [97], basically on the hypothesis that the main contribution to the resonance energy comes from the Coulomb potential of the electric interaction among the various protons of which the nucleus of Samarium consists. According to [97], after a certain amount of reasonable approximations, the energy bound translates into a bound on the variation of the electromagnetic coupling. A simple look at expression 4.2.11 shows that, in our scenario, the variation of this coupling over the time interval under consideration violates the Oklo bound. This bound seems therefore to rule out our theoretical framework. However, things are not so simple: the derivation of a bound on a coupling out of a bound on energies works much differently in our framework, and we cannot simply use for our purpose the results of [97]. Indeed, in our framework what varies with time is not only the fine-structure constant, but also the nuclear force, and the proton and neutron mass as well. Of relevance for us is therefore not a bound on a coupling, derived under the hypothesis of keeping everything else fixed, but the bound on the energy itself [97]:

$$-0.12\,\mathrm{eV} \, < \, \Delta E \, < \, 0.09\,\mathrm{eV}\,. \tag{5.4.6}$$

In order to give an estimate of the amount of the energy variation over time, as expected in our framework, we don't need to know the details of the evaluation of the resonance energy starting from the fundamental parameters of the theory. To this purpose, it is enough to consider that, whatever the expression of this energy is, it must be built out of i) masses, ii) couplings (electro-weak and strong) and iii) the true fundamental constants (the speed of light c, the Planck constant $\hbar$, and the Planck mass $\mathrm{M_p}$). Working in units in which the latter are set to 1 (reduced Planck units), all parameters of points i) and ii) scale as a certain power of the age of the universe. As a consequence, the resonance energy itself mainly scales as a power of the age of the universe:

$$E \sim a\mathcal{T}^{-b} \,. \tag{5.4.7}$$

(More generically, it could be a polynomial: $E \sim a_1\mathcal{T}^{-b_1} + a_2\mathcal{T}^{-b_2} + \ldots + a_n\mathcal{T}^{-b_n}$. In this case, to the purpose of checking the agreement with a bound, it is enough to look at the dominant term). We can fix the exponent b by comparing the expression, evaluated using the present-day age of the universe, with the value of the resonance, that we take from [97]:

$$E \sim a\mathcal{T}^{-b} = 0.0973\,\mathrm{eV} \times 1.2 \times 10^{-28} = 1.2 \times 10^{-29}\,\mathrm{M_P} \,. \tag{5.4.8}$$

In order to solve the equation, we would need to know the coefficient a, something we don't. However, as long as we are just interested in a rough estimate, it is reasonable to assume that, since this coefficient mostly accounts for possible symmetry factors, it may affect the value of the result for about no more than one order of magnitude. Inserting the value $\mathcal{T} \sim 5 \times 10^{60}\mathrm{M_P^{-1}}$ for the age of the universe, we obtain:

$$b \sim \frac{1}{2} \,, \tag{5.4.9}$$

and finally:

$$|\Delta E| \sim \frac{1}{10}E \sim 0.01\,\mathrm{eV} \,. \tag{5.4.10}$$

over a time of two billion years. This is compatible with the Oklo bound, eq. 5.4.6.

From the Oklo data one tries also to derive a bound on the adimensional quantity

$$\beta \equiv G_F m_\mathrm{p}^2 (c/\bar{h}^3) \,. \tag{5.4.11}$$

In this case, our discussion is easier, because we know the scaling of all the quantities involved [3]. Once again, we have to deal with a quantity that scales as a power of the age of the universe. At present time, we have:

$$\beta \sim \mathcal{T}^{-b_\beta} = 1.03 \times 10^{-5} \,. \tag{5.4.12}$$

Inserting the actual value of the age of the universe, we obtain $b_\beta \sim \frac{1}{12}$. Over a time interval of around 1/5 of the age of the universe, this gives a relative variation:

$$\frac{\Delta\beta}{\beta} \sim 0.017 \,, \tag{5.4.13}$$

to be compared with the one quoted in ref. [97]:

$$\frac{|\beta^{\mathrm{Oklo}} - \beta^{\mathrm{now}}|}{\beta} < 0.02 \,. \tag{5.4.14}$$

Both results 5.4.10 and 5.4.13, although still within the allowed range of values, seem to be quite close to the threshold, beyond which the model is ruled out. One would therefore think that a slight refinement on the measurement and derivation of these bounds could in a near future decide whether it is still acceptable or definitely ruled out. Things are not like that. Indeed, as we already stressed in several similar cases, the *entire* derivation of bounds and constraints, involving at any level various assumptions about the history of the universe and therefore of its fundamental parameters, should be rediscussed within the new theoretical framework: it doesn't make much sense to compare pieces of an argument, extracted from an analysis carried out in a different theoretical framework, with different phenomenological implications. To be explicit, in the case of the derivation of the Oklo bounds, one should reconsider the entire derivation of absorption thresholds and resonances. We should therefore better take into account from the beginning the time variation of all masses, and in particular the neutron and proton masses, as well as couplings. Perhaps a more

[3] We recall that $G_F/\sqrt{2} = g^2/8M_W^2$. Therefore, $\beta = \pi\alpha m_{\mathrm{p}}^2/\sqrt{2}M_W^2$. For times much higher than 1 in reduced Planck units, the proton mass can be assumed to scale approximately like the mean mass scale 4.3.26.

meaningful quantity is then not anymore the pure resonance shift, but this quantity rescaled by the neutron mass. In this case, the effective variation of interest for our test is not 5.4.10, but:

$$\frac{\Delta(E/m_{\rm n})}{(E/m_{\rm n})} \approx \frac{\Delta\mathcal{T}^{-\frac{1}{9}}}{\mathcal{T}^{-\frac{1}{9}}} \sim 0.02\,, \tag{5.4.15}$$

a variation one order of magnitude smaller than 5.4.10 ($\Delta E/E \sim 0.1$). Analogous considerations apply also to the case of the second bound 5.4.13, basically equivalent to the nucleosynthesis bound.

5.4.0.3 The nucleosynthesis bound

Bounds derived from nucleosynthesis models are even more questionable: they certainly make sense within a certain cosmological model, but, precisely because of that, they cannot be simply translated into a framework implying a rather different cosmological scenario. Once again, the only anchor points on which we can rely are the few "pure" experimental observations, to be interpreted in a consistent way in the light of a different theory. The point of nucleosynthesis is that there is a very narrow "window" of favourable conditions under which, out of the initial hot plasma, our universe, with the known matter content, has been formed. Of interest for us is the very stringent condition about the temperature (and age of the universe) at which the amount of neutrons in baryonic matter have been fixed. As soon as, owing to a cooling down of the temperature, the weak interactions are no more at equilibrium, the probability for a proton to transform into a neutron is suppressed with respect to the probability of a neutron to decay into a proton. Owing to their short life time, comparable to the age of the universe at which the equilibrium is broken, basically almost all neutrons rapidly decay into protons, except for those that bound into ^{4}He. Nucleosynthesis predicts a fraction of ^{4}Helium and Hydrogen baryon numbers ($\sim 1/4$) in the primordial universe, which is in good agreement with experimental observations. The formula for the equilibrium of neutron/proton transitions is given by:

$$\frac{\mathrm{n}}{\mathrm{p}} = \mathrm{e}^{-\Delta m / kT} \sim 1 \,, \tag{5.4.16}$$

where $\Delta m = m_{\mathrm{n}} - m_{\mathrm{p}}$. In the standard scenario, this mass difference is a constant, and the temperature runs as the inverse of the age of the universe. The equilibrium is broken at a temperature of around 0.8 MeV, when (n/p) $\simeq 1/7$. In our framework too the temperature runs as the inverse of the age of the universe, but the mass difference Δm is not a constant: all masses run with time. At large times ($\mathcal{T} \gg 1$ in Planck units), we are in a regime in which we can use the arguments of section 4.3.6, in order to conclude that, being the u and d quark masses much lighter than the neutron mass scale, we can consider Δm as a perturbation of $m \simeq m_{\mathrm{n}}$. In this regime, the neutron-proton mass difference is basically of the order of the constituent quarks mass difference, and we have reasons to expect that it also runs accordingly. It would therefore seem that, in our case, going backwards in time, the ratio (n/p) remains lower than in the standard case, and the equilibrium 5.4.16 is attained at a temperature much higher. However, to the purpose of determining the processes of the nucleosynthesis, essential is not just the scaling of the equilibrium law of the neutron-to-proton ratio, but also that of the mean life of the neutron. It is the combined effect of these two quantities what determines the primordial baryon composition. In the usual approach, the neutron mean life is assumed to be constant. Being related to the neutron decay amplitude, i.e. to the volume occupied by the neutron in the phase space, in our framework this quantity too is not constant. In order to see what in practice changes in our scenario with respect to the standard one, instead of attempting to guess what the scaling behaviour of the neutron mean life could be, we can proceed by considering, instead of the pure running of the equilibrium equation, the *reduced running at fixed neutron mean life*. Certainly the mean life is constant if the neutron mass is constant. The quantity of interest for us is therefore the scaling of the mass difference, as measured in units of the neutron mass itself. According to our considerations of above, we have:

$$\Delta m_{\rm red}(\mathcal{T}) \equiv \frac{\Delta m}{m_{\rm n}} \sim \frac{\mathcal{T}^{p_{(u-d)}}}{\mathcal{T}^{p_{\rm n}}}, \tag{5.4.17}$$

where $p_{(u-d)}$ and $p_{\rm n}$ are exponents corresponding to the up-down quark mass difference and to the neutron mass respectively. This running is expected to hold not only at present time but also at a temperature of ~ 1 MeV, which is anyway much lower than the Planck scale. We can therefore compare our prediction with the standard one by simply considering the relative deviation of equation 5.4.16 from its standard value, as obtained by replacing the constant mass difference Δm with $\Delta m_{\rm red}(\mathcal{T})$:

$$\frac{\rm n}{\rm p} = {\rm e}^{-\Delta m/kT} \quad \to \quad \left(\frac{\rm n}{\rm p}\right)_{\rm red} \equiv {\rm e}^{-\bar{m}_{\rm n}\Delta m_{\rm red}(\mathcal{T})/kT}, \tag{5.4.18}$$

where $\bar{m}_{\rm n}$ is the *fixed*, time-independent present-day value of the neutron mass. Therefore, in the standard case $({\rm n/p})_{\rm red}$ coincides with (n/p). According to the mass values given in section 4.2, we have:

$$\Delta m_{\rm red}(\mathcal{T}) \approx \mathcal{T}^{-\frac{1}{24}}. \tag{5.4.19}$$

Considering that the time variation between the point $\mathcal{T}_f$ of the breaking of equilibrium and the present day is of the order of the age of the universe itself, $\Delta T \equiv \mathcal{T} - \mathcal{T}_f \sim \mathcal{T}$, we obtain approximately that the integral variation of $x \equiv \Delta m_{\rm red}(\mathcal{T})$ over this time interval is:

$$\Delta x \sim \frac{1}{24} x. \tag{5.4.20}$$

The "reduced value" of (n/p), $({\rm n/p})_{\rm red}$, is now modified to:

$$\left(\frac{\rm n}{\rm p}\right)_{\rm red.} : \frac{1}{7} \to \sim \frac{1}{7}\left(1 - \frac{\ln 7}{24}\right) \approx 0.131. \tag{5.4.21}$$

This value leads to a ratio X_4 of helium to Hydrogen of around:

$$X_4 \sim 0.232, \tag{5.4.22}$$

still in excellent agreement with what expected on the basis of today's most precise determinations (for a list of results and references, see ref. [63]).

6 The phases of the natural evolution

Palaeontological observations seem to indicate that the evolution of life would not occur as a smooth, continuous progression, but would be characterized by relatively short periods of "sudden" mutation, separated by longer, more or less stable periods. For instance, it has been observed that the species of hominids, from primates to Homo sapiens, is characterized by an evolution toward an increasing craniofacial contraction, which makes possible an expansion of the volume of the brain, and appears to take place at specific periods in which a big step forward is made, followed by longer periods in which this kind of mutagenesis seems to be "at rest" [98]. This progressing through "steps" seems in some way to call into question certain aspects of the (neo-)Darwinian theory of the evolution through natural selection. Why should not all the possible directions, i.e. all the possible mutations, be statistically generated at the same time? Why should then evolution not be a continuous process? This has even induced to talk about "ontogenesis" for this kind of mutations, and mathematical models have been investigated, in order to explain this behaviour [99, 100, 101].

Of interest for us is here the biophysical dynamics of evolution, which seems to occur through a sequence of steps forward and rests, and this not only with regard to the human species, but also more in general to the big Eras of life on the Earth. In this chapter we discuss how this fits within our theoretical scenario. We have seen that, during the cosmological evolution, all fundamental mass scales m_i, as well as the couplings of elementary particles α_j, run as appropriate roots of the (inverse) age of the universe. Although complex systems (atoms, molecules) consist of several elementary particles bound together in a complicated way, so that their mass scale is not just given by the product of masses and couplings running as powers of the age of the universe, since *every* such element has a power-law scaling, also the mass and the energies of complex systems can be expressed as a sum of powers of the age of the universe. On the large scale there is therefore a dominant behaviour, which can always be reduced, up to normalization coefficients, to a power-law dependence on the age of the universe:

$$E_p \sim \frac{1}{\mathcal{T}^{1/p}} + \mathcal{O}\left(\frac{1}{\mathcal{T}^{1/q}}\right), \quad p > q > 1. \tag{6.0.1}$$

At our present time, the rate of variation of couplings, masses, and energies, is very small, irrelevant for our experience of every day. However, it becomes significant as seen on a cosmological scale. But its effect is appreciable also at “intermediate” scales, such as those of the evolution of life, where it can show out in “fine-tuning” effects. Among these are precisely the cases of natural evolution we are going to discuss. Here we will discuss how the sequence of these evolutionary steps, as well as the relatively short duration of the intervals of “rapid” progress of the evolution, can be explained entirely within the laws of molecular physics and the Darwinian theory of natural evolution.

6.1 The evolution of Primates

Let’s consider first the example referring to the most recent series of evolutionary mutations: the evolution of primates along steps of increasing cranio-facial contraction, summarized in figure 6.1. It is clear that the duration of these periods increases as we go back in time to earlier ages, although no simple mathematical relation seems to relate them. Once expressed in units of the age of the universe, the periods $\mathcal{T}_n$ of the primates-to-human history show a behaviour much less unfamiliar. Indeed, as we are going to see, they approximately arrange into a power series:

$$\mathcal{T}_n \approx k\, n^q, \tag{6.1.1}$$

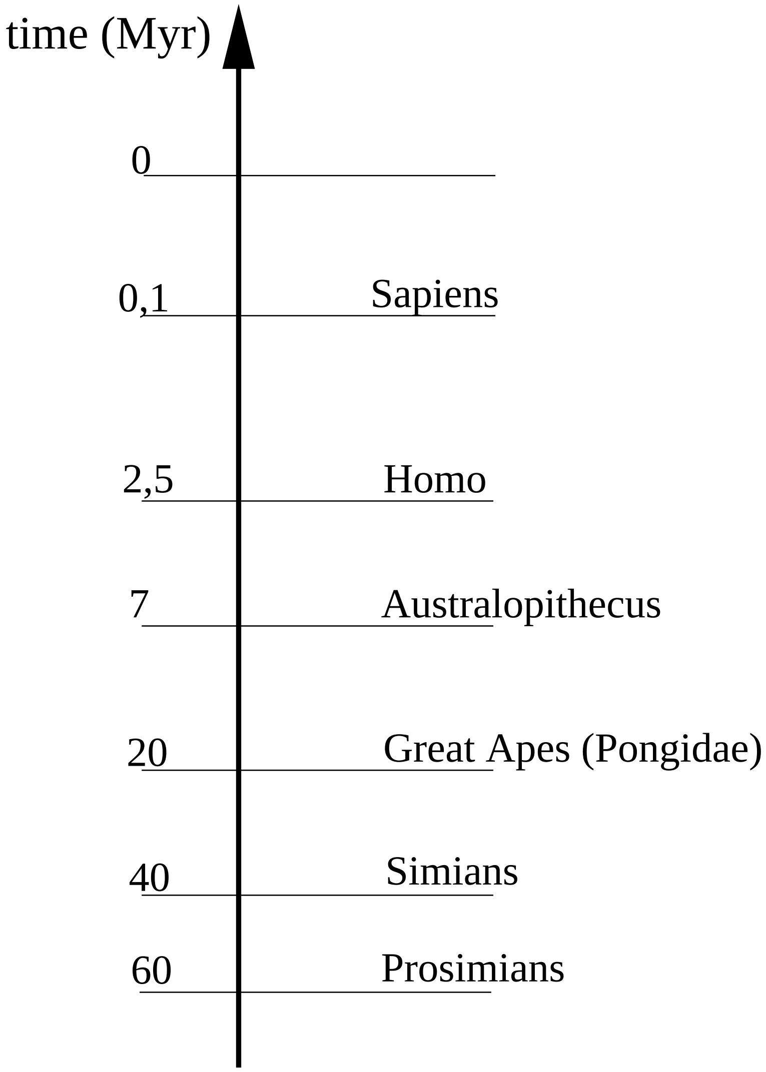


Figure 6.1: The steps of increasing cranio-facial contraction of hominids, according to ref. [98], as measured in millions of years.

for some positive numbers k and q, $0 < q < 1$, and n running on the natural numbers. What produces this behaviour? The fact that mutations seem to occur during a very short time, as compared to the duration of the stable phases, recalls the typical width of a resonance threshold in energy absorption processes. In a quantum system, energy levels are quantized and in general discrete; this is true at least as long as we consider a bound system and its binding energies, a situation to which the DNA corresponds with good approximation. Mutagenesis is a process produced by a change in the DNA structure. At the molecular level, what happens is that, as a consequence of the absorption of a certain amount of energy (e.g. radiation of a certain frequency), protons and/or electrons "jump" to different positions, and form new bonds. Let's consider to expose the DNA to a certain kind of radiation. The energy that hits the probe is quantized, and is related to the frequency ν, or the wavelength λ, of the radiation, according to the Compton law:

$$E_{\text{source}} = h\nu = \frac{c}{\lambda} \,. \tag{6.1.2}$$

Also the energy levels of the target molecule are expected to be quantized. The typical energies of mutagenetic processes are the object of several investigations, based on approximations of the DNA sequence as a crystal, or in general a system bound in a certain region [102, 103, 104, 105]. In general, the absorption spectrum is discrete:

$$E_{DNA} = \{E(n)\} \,, \qquad E(n) = k_n E_0 \,, \tag{6.1.3}$$

where k_n is a certain coefficient and n runs on (a subset of) the natural numbers. The radiation energy 6.1.2 can be absorbed by the DNA molecule, and produce a change in its structure, only if it corresponds to one of the discrete levels of its spectrum. In this case, we have a resonance of the absorption probability:

$$E_{\text{source}}|_{\text{res.}} \cong E(n)_{\text{target}} \,. \tag{6.1.4}$$

A series of evolutionary steps, such as those of the progressive craniofacial contraction, corresponds to a specific change of the DNA structure, possibly induced by a change of one or more proton bonds, that could be a transition of the kind considered in ref. [102], or something similar. Which molecular bonds do correspond to a certain degree of contraction is not known. However, it is not unreasonable to think that the amount of contraction is related to the number of bonds which underwent an "elementary" transition in the DNA molecule. Let's make the hypothesis that this is indeed the case. A larger degree of mutation would then correspond to a larger number of elementary transitions. In order to induce one such change, an "elementary step" A, the DNA molecule must absorb an energy:

$$E_A \;=\; E(n_A) \;=\; k_{n_A} E_0 \,, \tag{6.1.5}$$

for some quantum number $n = n_A$. Let's suppose that this is precisely induced by the absorption of energy coming from an external source of radiation. In order to induce the evolutionary mutation under consideration, we must therefore have:

$$E_{\rm source}|_{\rm res.} \;\cong\; E(n_A)_{\rm target} \,. \tag{6.1.6}$$

A discrete series of resonance points along the time axis is only possible if the two energy scales run as independent functions of time. The amount of change at any such point should be related to the time width of the resonance.

As anticipated, let's suppose mutations are induced by radiation. There are several candidates for a source of radiation able to induce genetic mutations: the UV radiation, mostly coming from solar light, the natural radioactivity, and cosmic rays. However, X and cosmic rays are extremely energetic, and the mutations they induce are in general not "evolutionary" but "destructive". The radiation that in practice can induce molecular changes leading to new forms of life, not just to the death of an organism, is the ultra-violet radiation, and perhaps an even less energetic one. Therefore, the energy spectrum of the source should basically be the one of the electronic transitions, giving rise to the known atomic emission spectra (in the case of hydrogen, the Lyman series etc...).

During the cosmological evolution, the spectrum and the amount of this type of radiation have changed, according to the evolution of the stars and in particular of the solar system. However, for what matters our problem, restricted to a very recent era of the evolution of the universe, it can be considered a sufficiently regular background [1]. Were the energy levels of the source, and of the target DNA, constant (as they are normally assumed to be), the mutation process would be progressive: the elementary transition would be constantly related to a certain spectral line, or a bunch of spectral lines. The rate of absorption would be proportional to the intensity of the source (almost constant), leading to a statistically continuous increase of the number of changed bonds in the DNA molecule. We would therefore observe a continuous evolution of primates. Since both the emitted radiation and the ground energy scale of the DNA bonds are functions of elementary energy scales and couplings, in our theoretical framework they have a dominant behaviour given as in 6.0.1. This means that, in first approximation, they run as two independent powers of the age of the universe:

$$E_{\text{source}} \approx \frac{k_{\text{s}}}{\mathcal{T}^{p_{\text{s}}}}\,; \tag{6.1.7}$$

$$E_{\text{target}} \approx \frac{k_{\text{t}}}{\mathcal{T}^{p_{\text{t}}}}\,, \tag{6.1.8}$$

where p_{s}, p_{t} are real numbers $0 < (p_{\text{s}}, p_{\text{t}}) < 1$, and k_{s}, k_{t} are coefficients that collect the contribution of symmetry factors and encode the dependence on the quantum numbers labelling the energy levels. At a generic time $\mathcal{T}$, the radiated energy doesn't correspond to any energy gap of the target. Let's suppose that at a certain age $\mathcal{T}_i$ we have a resonance with some spectral line of the source:

$$E(n_A) \approx E_{\text{source}}(n, m)\,, \tag{6.1.9}$$

[1] We refer here to the frequencies of the spectrum, and in general the cosmological running of the fundamental physical parameters. We don't consider variations due, for instance, to the solar activity, that don't affect such properties. We will comment about these effects in section 6.3.

where (n, m) is a shorthand notation that indicates the quantum numbers of the two energy levels involved in the transition producing the radiation in the source. When 6.1.9 is satisfied, energy can be absorbed, making possible for the system to undergo a class of genetic mutations, corresponding to new possible DNA molecular changes. Statistically, in a short time, corresponding to the width of the resonance, all possible mutations are tried out. There is therefore not necessarily a unique kind of mutation. The maximal transition probability is attained at the peak of the resonance. Out of this point, the probability rapidly decays, at a speed depending on the characteristic width of emission and absorption spectra. After the time window of the resonance, these transitions are no more possible (i.e. they are extremely suppressed), and the rate of the mutagenesis process drops down dramatically. Natural selection will then decide which one(s) among all the mutations will survive. The system will then "stabilize" until a new resonance threshold opens up. Suppose this was a facial bone contraction enabling a larger brain volume. We get a certain amount of contraction-inducing transitions (i.e. a certain amount of changed DNA bonds), depending on the width of the resonance window. Then the process stops till the new resonance. This occurs when the same kind of molecular transitions are induced by the next spectral line that turns out to meet the condition 6.1.9. If a larger brain is a mutation favoured by natural selection also at later times, then, at the next resonance time, Nature will favour again the same kind of transition; the suspended process of contraction will be resumed and progress for another while, leading to the birth of species of primates with a still larger brain.

We can give a rough estimate of the separation between subsequent resonance times. First of all, let's see what is the order of magnitude we should expect for the exponents $p_{\rm s}$ and $p_{\rm t}$ of eqs. 6.1.7 and 6.1.8. For the emission scale, under the hypothesis of an atomic origin of the radiation, whatever is the source of radiation in first approximation the atomic energy levels are given as some numbers multiplied by the Rydberg constant R. This is strictly true only in the simplest case, the hydrogen atom, in which case the energy levels are given by:

$$E_{\rm source}(n,m) \,=\, h\nu \,=\, R\left(\frac{1}{m^2}-\frac{1}{n^2}\right)\,, \tag{6.1.10}$$

where:

$$R \,\approx\, R_\infty = m_e\alpha^2/4\pi \quad (\times\, c/\hbar)\,, \tag{6.1.11}$$

where m_e is the electron mass and α the fine-structure constant (in our framework, neither of them is constant). The highest energy, ultraviolet series, is obtained with $m = 1$ (Lyman series). More in general, the energy levels have more complicated expressions, and, for heavy elements with many electrons, one has to consider also relativistic effects scaling as $m_e\alpha^4$. However, as long as we are interested in a rough estimate, we will assume here that the energy levels of our source have an effective approximate hydrogen-like spectrum. This hypothesis is on the other hand supported by the consideration that hydrogen is the most common element in the universe. We expect therefore that the energy levels behave approximately as the Lyman series:

$$E_{\rm source} \,\approx\, R_\infty\left(1-\frac{1}{n^2}\right)\,. \tag{6.1.12}$$

For the target DNA molecule, the energy levels of interest for us are not those corresponding to a transition among the positions of the electrons but those of the protons (see for instance refs. [102, 106]). A typical dominant term of DNA energies could then be something like $E_0^{\rm target} \,\approx\, m_{\rm p}\alpha^2$. Therefore, although we don't know the exact details of the system, we can reasonably assume that the DNA energy, $E_{\rm target}$, is above and runs slower than the energy of the cosmic source.

According to the results of chapter 4, both the electron mass and the electric charge (the fine-structure constant α) run as positive roots of the inverse of the age of the universe. This means that the Rydberg constant too scales as a certain root of the age of the universe. At sufficiently large times as compared to the Planck length (as is the case of the evolution of life), also the proton mass roughly scales as a root of the age of the universe. With reference to equations 6.1.7 and

6.1.8, we can therefore identify:

$$\frac{1}{\mathcal{T}^{p_{\rm s}}} \sim R_{\infty} = R_{\infty}(\mathcal{T}) \equiv E^{0}_{\rm source}(\mathcal{T})\,; \tag{6.1.13}$$
$$\frac{1}{\mathcal{T}^{p_{\rm t}}} \sim E^{0}_{\rm target}(\mathcal{T})\,. \tag{6.1.14}$$

The resonance condition 6.1.9 at a time $\mathcal{T}_i$ can be written as:

$$\mathcal{T}_i^{p_{\rm s}-p_{\rm t}} \approx k_{n_A} \times \left(1 - \frac{1}{n_i^2}\right)\,, \tag{6.1.15}$$

where, according to our previous considerations, $p_{\rm s} > p_{\rm t}$. Since as time goes by the energy scale of the source becomes smaller and smaller as compared to the DNA energy scale, subsequent matchings of energies occur by jumping to higher energy levels of the source, therefore toward higher n. Expression 6.1.15 neglects however the fact that, after the DNA sequence undergoes a step of mutation, its energy levels are no more the original ones: a different structure implies in general a different spectrum of energies. At this stage, we are not able to quantify this phenomenon, however we can expect that, since the natural evolution occurs towards a higher degree of complexity characterized by an overall higher energy, the new spectrum of energy levels consists in general of a finer pattern of absorption bands running at a higher scale. The new matching point with a higher energy level of the source should occur therefore earlier than what 6.1.15 predicts. As a matter of fact, this should lead to a shortening of the time elapse between subsequent resonance points, as compared to what predicted by 6.1.15. Moreover, as long as the density of energy levels increases, we must expect also an increase of the probability of overlapping of subsequent resonances. After a certain time, the progression of discrete steps should then effectively "saturate" to a continuum.

We will test hour hypothesis by working out the sequence of time intervals by solving the equation 6.1.15 for $n_i = n_{\rm s}$, $n_{i+1} = n_{\rm s} + 1$, $n_{i+2} = n_{\rm s} + 2, \ldots$, $n_{\rm s}$ being a typical point in the hydrogen series of the source. Let's introduce $q \equiv 1/(p_{\rm t} - p_{\rm s})$. Clearly, $q > 1$. We can then write equation 6.1.15 as:

$$\mathcal{T}_i \approx \left[k_{n_A} \times \left(1 - \frac{1}{n_i^2} \right) \right]^q . \tag{6.1.16}$$

In order to verify our hypothesis, we fit equation 6.1.16 over five points in the history of the universe, corresponding to the turning periods in which mutagenesis has produced the evolution of the human species from the Australopithecus to the Homo sapiens, illustrated in figure 6.1 of page 233. For the age of the universe, we use the value obtained in section 4.3.2.5, namely:

$$\mathcal{T} \; = \; 1.262028 \times 10^{10}\, yr \; = \; 5.038816 \times 10^{60} \mathrm{M}_{\mathrm{P}}^{-1} . \tag{6.1.17}$$

In order to get rid of big numbers and constant parameters, we plot therefore the quantity:

$$y(x) \equiv \frac{\mathcal{T}_{\ell+x}}{\mathcal{T}_\ell} , \tag{6.1.18}$$

for the five values from "Simians" to "Sapiens" as given in figure 6.1 [2]. From expression 6.1.16 we obtain:

$$y(N) \; = \; \frac{\mathcal{T}_{i+N}}{\mathcal{T}_i} \; \cong \; \left[\frac{1 - \frac{1}{(n_i+N)^2}}{1 - \frac{1}{n_i^2}} \right]^q . \tag{6.1.19}$$

For mass and energy scales ranging at present time from the meV to the keV scale, the exponents p_{s} and p_{t} have typical values in the range $\sim [\frac{3.5}{10}, \frac{3}{10}]$. Therefore, $q \gg 1$. Limiting the analysis to the first values of N, namely $N = 1, 2, 3, 4, 5, 6$, we can assume that $N \ll n_i$. Under these conditions, expression 6.1.19 can be approximated by:

$$y(N) \sim N^c , \qquad N = 1, 2, 3 \ldots \tag{6.1.20}$$

for some constant c. The small spacing of the periods, $\mathcal{T}_{i+1} - \mathcal{T}_i$, as compared to the age of the universe, tells us that $c \ll 1$. This

[2] We exclude the edge value corresponding to the Prosimians, on which we will comment later.

approximation is valid as long as we can write:

$$N \;\approx\; \left[\frac{1 - \frac{1}{(n_{\rm s} + \tilde{N})^2}}{1 - \frac{1}{n_{\rm s}^2}} \right]^{\frac{q}{c}} , \qquad \tilde{N} \equiv \pm \left(N - 1 \right) , \tag{6.1.21}$$

where we have shifted the value of N on the r.h.s. to $\tilde{N} = (N-1)$ in order to account for the fact that the point $N = 1$ of the interpolation corresponds to the point $\tilde{N} = 0$ on the r.h.s. For $n_{\rm s}$ sufficiently large we can expand the r.h.s. of 6.1.21 as:

$$\left[\frac{1 - \frac{1}{(n_{\rm s} + \tilde{N})^2}}{1 - \frac{1}{n_{\rm s}^2}} \right]^{\frac{q}{c}} \approx \left[1 + \frac{2\tilde{N}}{n_{\rm s}^3} + \mathcal{O} \left(\frac{1}{n_{\rm s}^2} \times \left(\frac{\tilde{N}}{n_{\rm s}} \right)^2 \right) \ldots \right]^{\frac{q}{c}} . \tag{6.1.22}$$

By keeping just the first two terms of the expansion, we have a binomial raised to the power q/c, and we obtain:

$$N \;\approx\; 1 + \left(\frac{q}{c} \right) \frac{2\tilde{N}}{n_{\rm s}^3} + \ldots , \tag{6.1.23}$$

where the neglected terms receive a contribution from what we neglected in 6.1.22, of order:

$$\sim \; \mathcal{O} \left[\left(\frac{\tilde{N}}{n_{\rm s}^2} \right)^2 \right] ; \tag{6.1.24}$$

and from the higher order terms in the binomial expansion:

$$\sim \; \mathcal{O} \left[\left(\frac{2\tilde{N}}{n_{\rm s}^3} \right)^2 \right] . \tag{6.1.25}$$

Equation 6.1.23 is approximately solved by:

$$n_{\rm s} \;\sim\; \left(\frac{2q}{c} \right)^{1/3} . \tag{6.1.26}$$

Notice that this kind of approximation may work also for atomic sequences other than the Lyman series. For a generic $1/m$ in expression 6.1.10 we would obtain an expression analogous to 6.1.23, simply with rescaled quantities: $n \to n/m$, $\tilde{N} \to \tilde{N}/m$, resulting in a solution:

$$n_{\mathrm{s}} \sim \left(\frac{2m^2 q}{c} \right)^{1/3} . \tag{6.1.27}$$

Therefore, we don't really need to assume that the energies of the source correspond to the Lyman series.

A similar behaviour is obtained also if we consider that the energy jump occurs between the energy levels of the target. In fact, the power-law behaviour 6.1.20 is basically due to the power-law scaling of the ratio of the basic scales $E^0_{\mathrm{source}}/E^0_{\mathrm{target}}$, and the fact that within a certain range the quantum energy levels can be approximated by a simple harmonic oscillator-like expression $E(n) \approx nE_0$. A quantum system in a box approximately corresponds to a three-dimensional harmonic oscillator. In the case of DNA, we can suppose that it roughly corresponds to a composite system of many harmonic oscillators. In this way, at the first order the coefficient k_n in 6.1.3 should be given by:

$$k_n \approx (n_{\mathrm{t}} + n_0) k_0 \,, \tag{6.1.28}$$

where k_0 is a scaling factor and n_0 is the ground energy, a quantum Casimir effect that, if in the case of a one-dimensional harmonic oscillator is $1/2$, in a complex system consisting of many harmonic oscillators can be a much larger number. If this is the case, then, keeping fixed the quantum numbers of the energy of the source, a power-law sequence like 6.1.20 is obtained as long as we can approximate:

$$\left(\frac{n_{\mathrm{t}} + \tilde{N} + n_0}{n_{\mathrm{t}} + n_0} \right)^{\frac{q}{c}} \approx 1 + \left(\frac{q}{c} \right) \frac{\tilde{N}}{n_{\mathrm{t}} + n_0} + \mathcal{O} \left(\frac{\tilde{N}}{n_{\mathrm{t}} + n_0} \right)^2 , \tag{6.1.29}$$

by retaining only the first two terms, and identifying this time:

$$\frac{q}{c} \sim n_{\mathrm{t}} + n_0 \,, \tag{6.1.30}$$

for some n_t. This is certainly possible, if the ground number n_0 is sufficiently large. In practice, the fact of having a sequence of the type 6.1.20 is related to the possibility of making a linear approximation of the spacing of the energy levels, either of the source or of the target, or both of them, into steps of equal separation, at fixed fundamental energy scale. Once the running of the latter is taken into account, this translates into a series of the type 6.1.1.

For what we have just discussed, it is reasonable to fit the ratios 6.1.19, referred to the five last steps of the evolution of primates, with the curve:

$$y \,=\, a\, x^c \,. \tag{6.1.31}$$

To stay more general, what we indeed fit is the curve:

$$y \,=\, a(x-b)^c \,. \tag{6.1.32}$$

By consistency, we expect to find a fit for $b \ll 1$. Using the value 6.1.17 for the age of the universe, the values $\mathcal{T}_i$ of the time periods illustrated in figure 6.1 can be approximated by:

$$\begin{aligned}
\mathcal{T}_1 &\approx 1.256028 \times 10^{10}\,\mathrm{yr}\,; \\
\mathcal{T}_2 &\approx 1.258028 \times 10^{10}\,\mathrm{yr}\,; \\
\mathcal{T}_3 &\approx 1.260028 \times 10^{10}\,\mathrm{yr}\,; \\
\mathcal{T}_4 &\approx 1.261328 \times 10^{10}\,\mathrm{yr}\,; \\
\mathcal{T}_5 &\approx 1.261778 \times 10^{10}\,\mathrm{yr}\,; \\
\mathcal{T}_6 &\approx 1.262018 \times 10^{10}\,\mathrm{yr}\,; \\
\mathcal{T}_7 &\approx 1.262028 \times 10^{10}\,\mathrm{yr}\,.
\end{aligned} \tag{6.1.33}$$

Their ratios are therefore:

$$\begin{aligned}
y(1) &= \mathcal{T}_2/\mathcal{T}_1 \approx 1.001592\,; \\
y(2) &= \mathcal{T}_3/\mathcal{T}_1 \approx 1.003185\,; \\
y(3) &= \mathcal{T}_4/\mathcal{T}_1 \approx 1.004220\,; \\
y(4) &= \mathcal{T}_5/\mathcal{T}_1 \approx 1.004578\,; \\
y(5) &= \mathcal{T}_6/\mathcal{T}_1 \approx 1.004769\,; \\
y(6) &= \mathcal{T}_7/\mathcal{T}_1 \approx 1.004777\,.
\end{aligned} \tag{6.1.34}$$

Consistently with our previous discussion, better than looking for an overall good fit, we try to see for which value of the parameters the first experimental steps are reproduced: later steps should be the more and more affected by the effects of the change in the DNA lattice structure, and by the thickening of the band spectrum in the radiating source. The results are shown in figure 6.2. Indeed, the agreement is obtained for:

$$\begin{aligned} a &= 1.001623372\,; \\ b &= 0.020001238\,; \\ c &= 0.002345118 \times 10^{-3}\,. \end{aligned} \tag{6.1.35}$$

These values are consistent with our hypothesis of having a close to 1, b and $c \ll 1$. As one can see, in the last steps the disagreement between experimental observations and the interpolation curve increases. Indeed, it seems that our present time falls in a phase in which the sequence of steps is close to saturation of the discrete series, at the transition to a phase of overlapping resonances. A higher density of energy levels implies that the series progresses toward configurations that get closer and closer to each other. We expect this to imply that also the changes induced by mutagenesis become the more and more frequent and small.

6.2 The great Eras of life: the Paleozoic, Mesozoic and Cenozoic steps

We expect a similar mechanism to be at the ground of evolutionary processes that don't refer only to the primates but to any form of life. A problem is to identify which sets of mutations can be grouped into classes corresponding to the same "basic" transition, and therefore can be arranged along the same series of neighbouring resonances. We can imagine that the evolutionary processes can be distinguished into several classes, according to the kind of molecular transitions they are controlled by. For instance, by looking at figure 6.3, one can figure out that the big subdivision into Paleozoic, Mesozoic and Cenozoic Eras of the natural evolution should not mix with the "sub-eras", the

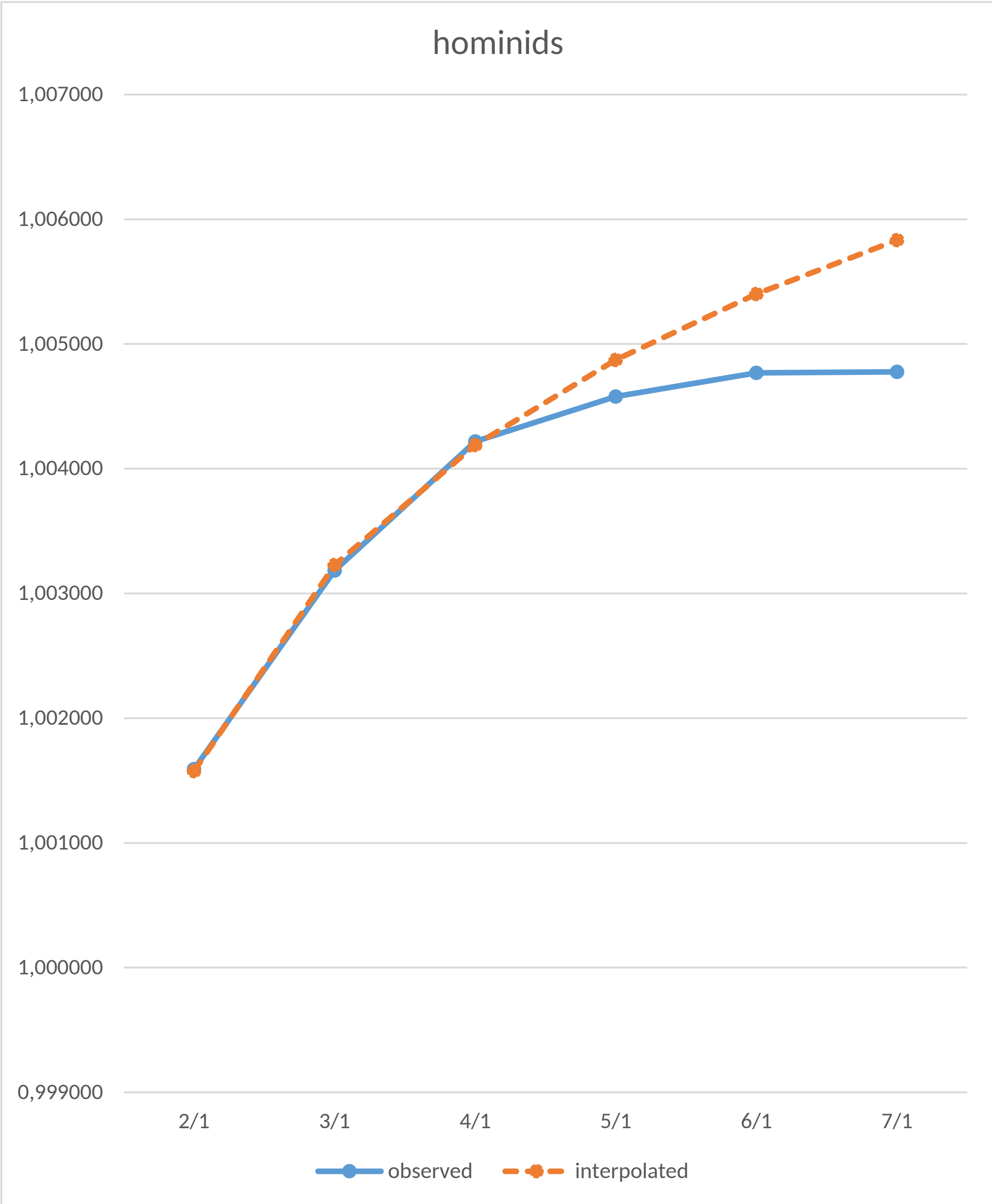


Figure 6.2: The steps of evolution of hominids. On the Y axis are reported the ratios of time periods, as derived from 6.1.33 and given in 6.1.34. The steps on the X axis range from the $Simians(2)/Prosimians(1)$ to $Present\ Time(7)/Prosimians(1)$.

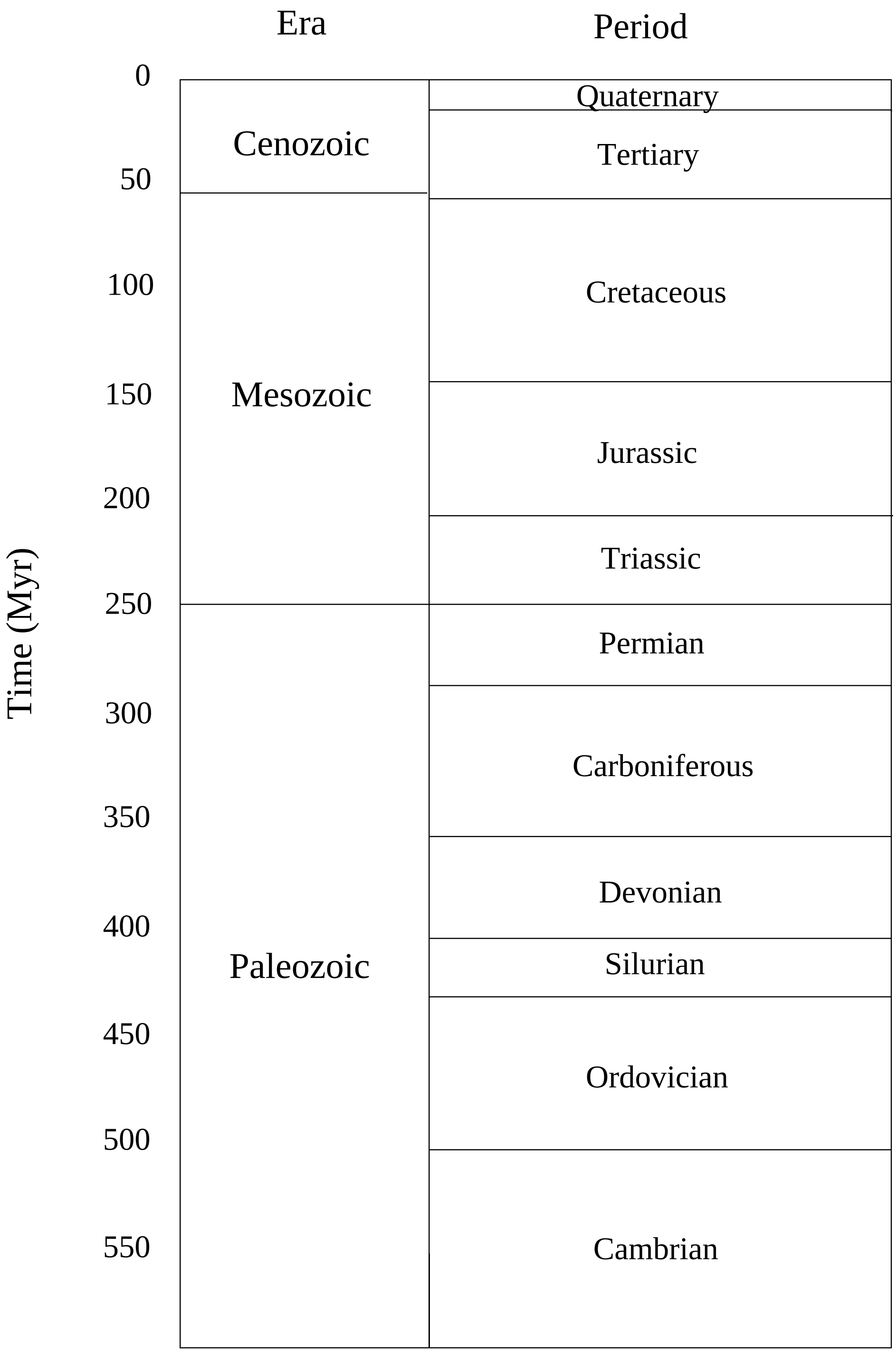


Figure 6.3: The great Eras of the evolution of life.

Periods such as the Triassic, the Jurassic etc..., although these periods not necessarily fit into subclasses of the main class of transition. This means that not necessarily "Triassic, Jurassic and Cretaceous" belong to the same main class, distinguished from the class formed by the set "Cambrian, Ordovician, Silurian, Devonian, Carboniferous, Permian". The beginning of the first era, the Paleozoic Era, is the time when most of the major groups of animals first appear in the fossil record, and is sometimes called the "Cambrian Explosion", because of the relatively short time over which this diversity of forms appeared. The Triassic-Permian extinction event too is something that took place in a relative short interval of time. Lastly, the end of the Mesozoic era is characterized by the sudden disappearance of dinosaurs. These facts strongly suggest that also the beginning and the end of these eras were marked by a rapid evolution, as due to the opening of new resonance thresholds allowing genetic mutation. We may ask whether also these big eras of the evolution of life roughly follow a power-law sequence.

In order to perform a similar analysis for the Paleozoic-Mesozoic-Cenozoic sequence, we need at least four time points, out of which to derive three ratios. From figure 6.3 we see that we are forced to include also our present time. This is somehow even more questionable than in the previous analysis, because we don't know whether we are now at the turning point of a new era. Our analysis will therefore be affected by even larger uncertainties than our previous was. Nevertheless, let us consider the transition times given by the beginning of Paleozoic (541×10^6 yr from present time), Mesozoic (252×10^6 yr from present time), Cenozoic (66×10^6 yr from present time). When expressed in terms of the age of the universe, we obtain:

$$\begin{aligned} \mathcal{T}_1 &\approx 1.207928 \times 10^{10}\,\mathrm{yr}\,; \\ \mathcal{T}_2 &\approx 1.236828 \times 10^{10}\,\mathrm{yr}\,; \\ \mathcal{T}_3 &\approx 1.255428 \times 10^{10}\,\mathrm{yr}\,; \\ \mathcal{T}_3 &\approx 1.262018 \times 10^{10}\,\mathrm{yr}\,, \end{aligned} \tag{6.2.1}$$

where we have added as fourth time the age of the universe, to account for our present time. From these we obtain the ratios:

$$\begin{aligned} \mathcal{T}_{2/1} &= 1.023925 \times 10^{10}\,\mathrm{yr}\,; \\ \mathcal{T}_{3/1} &= 1.039324 \times 10^{10}\,\mathrm{yr}\,; \\ \mathcal{T}_{4/1} &= 1.044787 \times 10^{10}\,\mathrm{yr}\,. \end{aligned} \tag{6.2.2}$$

By proceeding in the same way as in section 6.1, we plot the values $y(x)$. The coefficient of the curve 6.1.31 are now:

$$\begin{aligned} a &= 1.024393617\,; \\ b &= 0.021372812\,; \\ c &= 0.021204637\,, \end{aligned} \tag{6.2.3}$$

The curve is plotted in figure 6.2. Although the values of the interpolation coefficients are only approximately indicative, it is remarkable that the coefficient c differs from the c of section 6.1 by one order of magnitude. This value is higher than the statistical uncertainty due to the artifacts of the interpolation algorithm. The difference between the two coefficients is therefore something real, and signals that we are in the presence of absorption resonances corresponding to a different series and power law. This on the other hand is precisely what we should expect from genetic mutations of another kind: in this case, they would in general correspond to different DNA transition energies.

Our analysis suggests that also the three big eras of the evolution, the Paleozoic, Mesozoic and Cenozoic, are compatible with the interpretation as series of resonances. We may then ask whether the disappearance of dinosaurs, the event that marks the end of the Mesozoic era, could be ascribed to the appearance of more evolved competitors, perhaps coming from a mutation of already existing species. Palaeontological observations are in fact compatible with and extinction time that, although short when compared to the duration of an entire geologic era, amount to several thousand of years [3], a transition time perhaps longer than what we would have expected if it was produced by some "external" catastrophic event, and probably better suits to

[3]see for instance ref. [107].

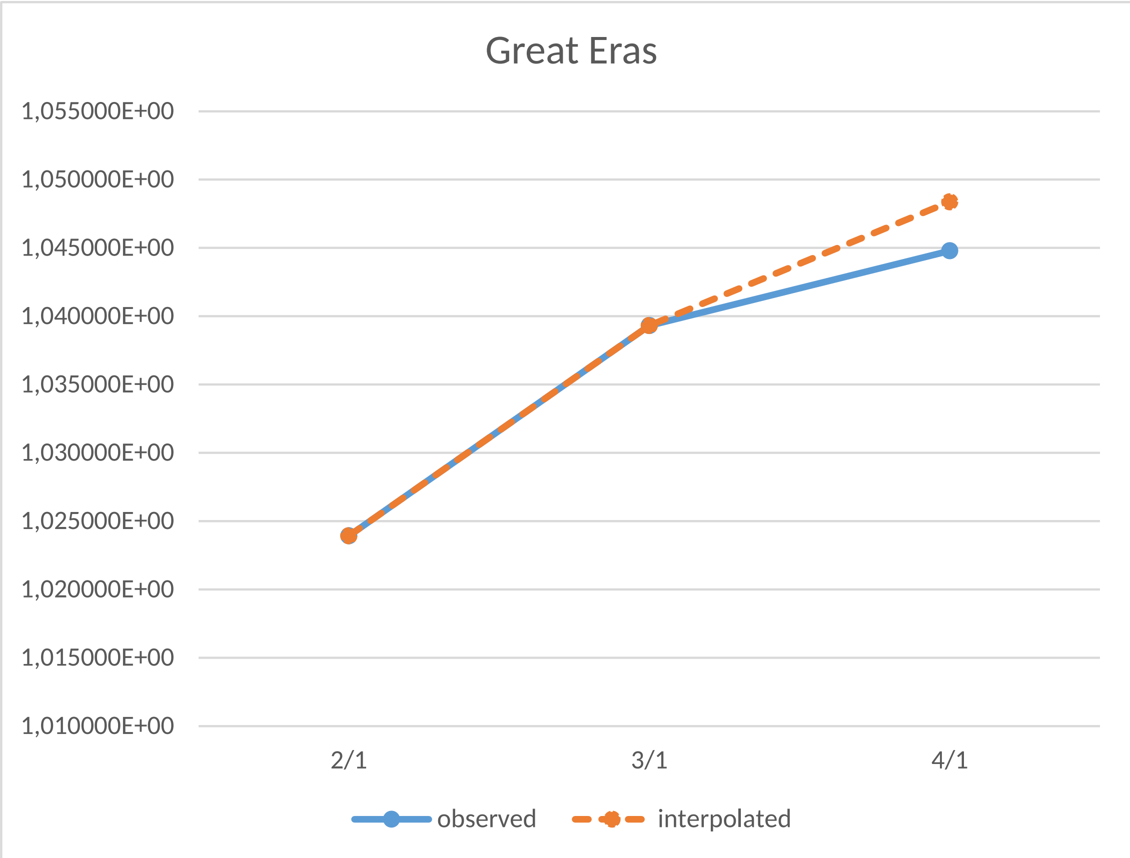


Figure 6.4: The ratios of time steps, observed and interpolated (dotted line), corresponding to figure 6.3.

a typical resonance width. We know that eventually mammals prevailed, although they already existed well before; could it be that a slight mutation finally gave them the necessary advantage to prevail over dinosaurs?

6.3 Remarks

At this point, several considerations are in order:

• Two different classes of the evolution, namely the one of the big eras of life on the earth, and the one of the primates, seem to arrange into sequences corresponding to DNA resonance energies. At our present state of knowledge, we cannot decide out of any doubt what distinguishes the sequence of the human evolution from the larger evolutionary scale of the three main eras of figure 6.3. In the case of the evolution of primates, we assumed that *the same* kind of molecular transition acts at any time there is a resonance condition. The amount of progress in the evolution, according to [98] proportional to the amount of cranio-facial contraction, would then be proportional to the number of occurred molecular transitions in the DNA. A priori it is not clear whether also in the case of the sequence of the big eras of figure 6.3, a unique kind of mutation is at work during all the turning periods. The seek for an answer could lead to a deeper understanding of the mechanisms of DNA transitions and their relation to the evolution.

• Obviously, different molecular transitions lead to different mutations. Therefore, the entire history of the evolution cannot fit into a single series. However, in general not necessarily all the steps of the evolution can be ordered into some series. A simple look at eras, ages and periods, shows that there are many "irregular" periods, which apparently cannot be arranged into any ordered sequence. Indeed, there can be a huge variety of combinations of DNA and source energy levels, leading to different mutations. Owing to the superposition of different mutations and different periods, the history of the evolution may not look so well ordered. It remains however a key point that these transitions occur at "discrete" points of the time axis, a feature that naturally fits with our scenario of time-running energy scales.

• The time spread of a mutation period does not depend only on the width of a resonance, but also on the fact that natural radiation is not "coherent", it has a certain spread of frequencies.

• The main source of UV radiation coming to the earth is the sun. Its activity is not constant; however, the solar phases involve the amount of produced radiation, not its being in resonance or not. As a consequence, under the hypothesis that the major cause of evolutionary mutagenesis is the solar light, what we expect is that variations of the solar activity affect the evolution process only if they fall within the time window of some resonance; in this case the mutation process can be accelerated (or slowed down).

• For simplicity, we did not consider mutagenesis of plants. In principle, these too could (should?) follow similar laws, and perhaps the full story about evolution of species is the result of an interaction/interference of all these phenomena.

All these considerations make sense only within a scenario in which, like in our case, the energy scales depend on time. Only in this case we obtain a discrete sequence of "resonance" periods. Otherwise, the full spectrum of emission from natural sources, as well as the complete spectrum of molecular energy levels, would be fixed and constant all along the history. The conditions for a genetic mutation would then be always the same, and mutations would be statistically generated without interruption. A step-wise progress of the evolution would then require completely different explanations.

When expressed in terms of the time separating these periods from our present time, as in figures 6.1 and 6.3, the power-law scaling, relation 6.1.1, is not explicit. The situation is reminiscent of the one of the law of a perfect gas, $PV = nRT$, in which the proportionality between pressure/volume and the temperature is only unveiled when the latter is expressed in terms of the absolute Kelvin scale. Analogously, here in order to see the relation we must express the time periods in terms of the absolute age of the universe.

Despite the high degree of approximation, our analysis suggests that the main steps are something "regular" and absolutely "programmed". Not by something external to the rules of natural evolution and selection; simply, something intrinsic of the fundamental laws of physics. The universe is expected to evolve toward more entropic configurations, in which the minimal energy step, which is also the size of the "unit cell" of the phase space, decreases. This agrees with the fact that the duration of the various phases decreases, making the more and more frequent the transition points. It however also means that the changes, the mutations, which are to be expected, should become less dramatic: more frequent, but also in the average smaller, steps.

7 High-temperature superconductivity

In this theoretical framework, the Heisenberg's uncertainty is the better satisfied as an equality the more is the geometry of the physical system "smooth". The energy uncertainty is bounded from below by the uncertainty of the most "classical" geometry, the one of a 3-sphere of radius $c\Delta t$, and energy content $\sim \left(M_{\mathrm{Pl}}^2 c^4/\hbar\right) \Delta t$ (this is also the basic geometry of the universe itself, see chapter 2). Indeed, in this theoretical scenario we can say that the fact of being also gravity quantized reflects in the fact that the degree of non-classicity of a physical system turns out to depend on its geometry, intended in the general relativistic sense of space distribution of energy. More complex quantum systems show a higher degree of quantum delocalization.

Quite a few physical systems look almost like a 3-sphere with almost the energy density of a black-hole. But, as long as we look at sufficiently extended bodies with a big mass, the uncertainty in position and momentum is negligible, and so are also the differences between the usual approach to quantum mechanics and our theoretical framework. As we move towards the atomic, and subatomic, scale, the difference between our theoretical framework and the usual approach becomes relevant. Superconductors are typical systems in which this phenomenon becomes critical and evident. These are materials that, although in themselves can be of huge extension, we are going to probe in their small scale properties. And, the more, by probes, the electrons, in a rather non-classical regime, such as the collective Cooper-pairs wave functions close to their ground energy. That is, where energy-momentum/time-position uncertainties play a relevant role, and where therefore quantum gravity effects show up more evidently. Apparently, this contradicts the popular idea that quantum gravity should become relevant only at the Planck scale.

7.1 Quantum gravity and superconductors

The phenomenon of superconductivity is explained in its grounds as due to the formation of pairs of electrons, that, behaving thereby "collectively" as bosons, can fulfill a narrow band of energy obeying to Bose-Einstein statistics. In other words, there can be very many within such a narrow band, so to produce a non negligible electric current [108]. In the BCS argument, essential for the occurring of this process is the existence of an attractive potential, attributed to the phononic response of the atoms of metal under charge displacement due to a motion of the electrons. This produces an energy gap Δ, and it has been shown that at the critical temperature most of the electrons pairs lie in an energy range of order Δ, at an energy which lies a gap Δ above the Fermi energy. This at least in the most simple formulation of the theory. More complicated structures of superconductors require modifications of this simple model, and eventually also weaken the existence of an energy gap as an essential feature of superconductivity, because there are conditions under which superconductivity exists even without an energy gap. We will not consider here the details of these model modifications and adjustments, which, as in any attempt to describe real, complex physical systems, are somehow unavoidable. For what interests our present discussion, it is important to consider that, whatever the derivation and the approximation introduced in order to reproduce a physical model can be, superconductivity remains related to the existence of a sufficient amount of electrons possessing a sufficient degree of non-locality. We can summarize this by introducing a "critical length" ξ which, for reasons that will become clear in the following, does not necessarily coincide with the "coherence path" it is usually talked about in the literature about superconductivity. For the time being, let us just assume that, according to a certain mechanism, which can reasonably be the one of phonon response of the BCS approach, Cooper's pairs do form and collect to a characteristic length higher than ξ when the temperature is sufficiently low. It must be stressed that the distribution of electrons is not a mathematical step function. Step functions are useful approximations introduced in practical computations. In reality, it is a matter of statistics. Therefore, one should never forget that, at any temperature, there will be a certain amount of pairs with typical length below ξ, and a certain amount above ξ. The relative amounts are a matter of temperature. If we call n the total number of electrons and n_S the number of electrons which are paired and with typical length larger than ξ, we can define the critical temperature T_c independently on the possible existence of an energy gap, just as the temperature at which $n_S/n \geq (n_S/n)_0$, $(n_S/n)_0$ being a certain well defined ratio, which does not need to be better specified. ξ is therefore a mean quantity. In traditional quantum mechanics, where gravity is switched off, ξ can only increase as a consequence of a higher localization in the space of momenta: $\langle \xi \rangle \sim \hbar/\Delta p \sim \hbar v_F/\Delta E$. In our quantum gravity scenario, $\langle \xi \rangle$ depends instead also on the complexity of the geometry of the system. We want to see how this comes about, and how, as a consequence, the critical temperature too will turn out to depend on the complexity of the geometry.

Let us consider the sum 2.1.16. It describes a universe "on shell"; namely, the universe "as it is". This means that there is no isolated system (particle or complex system of any kind), i.e. not embedded in its environment. In particular, there is no isolated system existing in a flat space. Not only the dominant geometry always contains the ground curvature of the universe, but any geometry ψ involved in the sum 2.1.16 is a distribution of $E = \mathcal{T}$ total energy degrees of freedom along a target space. Any geometry describes therefore a "whole universe". If we want to consider just a particular system, we must make an abstraction, and i) look at just a subset of the geometriess contributing to 2.1.16, ii) for any geometry of this subset, we must restrict our attention to a subregion of space. There is here a subtlety, because, in general, ψ does not describe a universe in three dimensions. As we said, this is true only for the dominant geometries. On the other hand, if it is true that the contribution of non-three dimensional, less entropic geometries is precisely what makes of the universe, and, in particular, of any subregion of it, a quantum system, it is also true that, in the concrete cases we want to consider here, a full bunch of these geometries, and precisely the most entropic ones, describe an energy distribution in a three dimensional space. Otherwise, we would not be able to talk of superconductors in the terms we are used to, namely, as well identified and (macroscopically) localized materials in a three-dimensional space. The physical systems we consider are therefore "at the border" between two descriptions: not anymore completely classical, but not even absolutely remote in the phase space, in order to completely escape the ordinary parameters of our perception of a three-dimensional space-time, and therefore of operational definition through a set of measurement and detection rules and experiments.

Let us concentrate our attention on just a small part the universe, a piece of superconducting material and, possibly, its close environment, with its atoms, electrons, magnetic fields etc., namely, all what constitutes our "experiment". Let us call $E^{(sc)}$ the energy of this portion of the universe. Of course, $E^{(sc)} < E$, the total energy of the universe (indeed, obviously $E^{(sc)} \lll E$). Let us consider the bunch of geometries of the universe that contain our superconductor, $\left\{\psi^{(sc)}\right\}$. Of course, for what we said *all* the geometries contributing to 2.1.16 do contain also the portion of universe in which our superconductor is placed. However, what we want to do here is to select the subset of geometries that contribute in a non-trivial way to form up the shape of the superconductor, not those that contribute, say, just for the ground curvature of space.

When we measure the energy of our experiment, the quantity that we detect is a mean value of energy, $\langle E^{(sc)} \rangle$, defined as [1]:

$$\langle E^{(sc)} \rangle \;=\; \frac{1}{\mathcal{Z}} \int_{\psi \in \left\{ \psi^{(sc)} \right\}} \mathcal{D}\psi \; \mathrm{e}^{S} \, E^{(sc)} \, . \tag{7.1.1}$$

Let us consider how energy can be distributed in the space, in order to form up our experiment of mean energy $\langle E^{(sc)} \rangle$. According to 2.1.16, the more a geometry is remote in the phase space, the less it weights in the sum out of which we should compute the mean total energy of the experiment. Since in a finite region of space we can arrange only a finite amount of energy (we can put at most one unit of energy per each unit of space, where units of energy are measured in terms of Planck mass, units of space in terms of cells of Planck length size), in order to get a certain amount of mean total energy $\langle E^{(sc)} \rangle$ we must sum up over a larger and larger number of geometries. The larger and larger, the more and more remote the average geometry we want to describe. Moreover, since in a finite region of space we can arrange only a finite number of different distributions of energy, as we go further with the remoteness, to sum up to the same fixed amount of local energy $\langle E^{(sc)} \rangle$ we must include geometries $\psi^{(sc)}$, in which $E^{(sc)}$ is supported in larger and larger space regions. In terms of traditional quantum mechanics, this means that the wavefunctions are more and more spread out in space.

As discussed in chapters 2 and 3, geometries can be classified according to the (finite) symmetry group of the distribution of energy degrees of freedom in the target space they correspond to. Their weight in 2.1.16 corresponds, by definition, to the number of times they occur in the phase space, in turn given by the number of equivalent ways they can be formed. The ratio of the weights in the phase space of two geometries can be expressed as:

$$\frac{W(\psi_i)}{W(\psi_j)} \;=\; \frac{||G_i||}{||G_j||} \, , \tag{7.1.2}$$

[1] All this can be put on a formal ground, by introducing an appropriate operator that, as is usual to do in the case of any generating functions, extracts from the logarithm of 2.1.16 the energy of a space domain around our experiment, but we don't want to bother here the reader with formalisms, rather to give the insight into the physical meaning of what we are doing.

where G_i and G_j are the symmetry groups, and $||G||$ indicates the volume of the group. This means that the more symmetric a geometry is, the higher is its weight in the phase space [2]. If a geometry ψ_j corresponds to a more broken symmetry group than a geometry ψ_i, it will be more remote, more "peripherical" in the phase space.

Let us introduce the concept of mean weight of our experiment, $W^{(sc)}$, and of mean volume of the symmetry, or volume $||G^{(sc)}_{\langle\psi^{(sc)}\rangle}||$ of the symmetry group of the mean geometry $\langle\psi^{(sc)}\rangle$ through:

$$\langle W^{(sc)}\rangle \;=\; \frac{1}{\mathcal{Z}} \int_{\psi\in\left\{\psi^{(sc)}\right\}} \mathcal{D}\psi \, \mathrm{e}^{S} \,, \tag{7.1.3}$$

and:

$$\frac{\langle||G^{(sc)}_i||\rangle}{\langle||G^{(sc)}_j||\rangle} \;=\; \frac{\langle W^{(sc)}_i\rangle}{\langle W^{(sc)}_j\rangle} \,. \tag{7.1.4}$$

Accordingly, we define the mean geometry $\langle\psi^{(sc)}\rangle$ as the geometry for which:

$$W\left(\langle\psi^{(sc)}\rangle\right) \;\stackrel{\mathrm{def}}{=}\; \langle W^{(sc)}\rangle \,. \tag{7.1.5}$$

Let us suppose we change the symmetry of the geometry of our superconductor, $\langle\psi^{(sc)}\rangle \to \langle\psi^{(sc)\prime}\rangle$, so that $||G^{(sc)}|| \to ||G^{(sc)\prime}|| = \frac{1}{2}||G^{(sc)}||$. In order to build up the same amount of energy $\langle E^{(sc)}\rangle$ we must consider geometries that correspond to the distribution of a higher amount of energy. How much more energy should we add, and how larger must be the space support? Approximately we must consider twice as much energy, implying that we must double the volume. Since superconductors are built-up in layers, the problem is basically one-dimensional. If we indicate with x the coordinate of the axis orthogonal to the two-dimensional layers, we have that, in order to maintain unchanged the value of $\langle E^{(sc)}\rangle$ and $||G^{(sc)\prime}||$, we need to consider distributions of energy of extension two times as large along the coordinate x. The

[2]Remember that the basic definition of space is discrete. Therefore, one work always with finite groups, for which $G_i \neq G_j \Leftrightarrow ||G_i|| \neq ||G_j||$ (see chapter 2).

linear spread in space of wavefunctions is therefore inversely proportional to the volume of the mean symmetry group:

$$\frac{\langle \Delta x \rangle}{\langle \Delta x \rangle'} = \frac{\langle ||G^{(sc)\prime}|| \rangle}{\langle ||G^{(sc)}|| \rangle} . \tag{7.1.6}$$

For any set of geometries with local symmetry group G_i, we may think of G_i as the little group of symmetry surviving after quotienting a larger group G through h_i. If we have two sets of geometries, ψ_i and ψ_j, obtained by quotientation from the same initial group: $G_i = G/h_i$, $G_j = G/h_j$, we have:

$$\frac{W(\psi_i)}{W(\psi_j)} = \frac{||h_j||}{||h_i||} . \tag{7.1.7}$$

Passing from the generic ψ_i, ψ_j to $\psi^{(sc)}$, $\psi^{(sc)\prime}$, and introducing correspondingly $h^{(sc)}$, $h^{(sc)\prime}$ instead of h_i, h_j, we can write 7.1.6 as:

$$\frac{\langle \Delta x \rangle}{\langle \Delta x \rangle'} = \frac{\langle ||h^{(sc)}|| \rangle}{\langle ||h^{(sc)\prime}|| \rangle} . \tag{7.1.8}$$

The relation 7.1.8 has been derived by imposing that the contribution of peripheral geometries to the mean energy $\langle E^{(sc)} \rangle$ remains constant. According to the definition and construction of the Heisenberg's uncertainty we gave in chapter 2, this relation tells us that $\langle ||h^{(sc)}|| \rangle$ can be viewed as an *effective* Planck constant:

$$\frac{\langle \Delta x \rangle \times \langle \Delta p \rangle}{\langle \Delta x \rangle' \times \langle \Delta p \rangle} = \frac{\langle ||h^{(sc)}|| \rangle}{\langle ||h^{(sc)\prime}|| \rangle} \equiv \frac{h_{\text{eff}}}{h'_{\text{eff}}} . \tag{7.1.9}$$

Up to an overall proportionality constant, which can be set to one, we can therefore write the quantum gravity version of the Heisenberg's uncertainty as:

$$\Delta x \Delta p \geq \frac{1}{2} \hbar_{\text{eff}} , \tag{7.1.10}$$

where $\hbar_{\text{eff}} \equiv h_{\text{eff}}/2\pi$ is related to the symmetry of a geometry through 7.1.9, 7.1.8, 7.1.6 and 7.1.4. Since increasing $||h||$ corresponds to increasing the complexity of the geometry, things work as if, by increasing the complexity of its structure, the system would become less and less classical, more and more quantum mechanical.

We have identified the critical temperature of superconductivity T_c as the temperature at which a well defined portion of electronic-bosonic states are delocalized at least as much as a critical length ξ. It is not necessary here to go into the details of the actual computation of T_c within a specific model. It is enough to know that it is obtained by integrating over a statistical distribution of states, and that the latter is expressed in terms of weights depending on E/kT. T can be viewed as the unit of measure of E: everything depends in fact on the ratio E/T, and a rescaling of E is compensated by a rescaling of the temperature T while keeping fixed the ratio E/T. Let us consider once again our example of the two geometries characterized respectively by $||G^{(sc)}||$ and $||G^{(sc)\prime}|| = \frac{1}{2}||G^{(sc)}||$. In the primed case, the same delocalization in space as in the unprimed geometry corresponds to one-half the unprimed energy. Since both energies are effectively "measured" in units of T, instead of talking of half energy, we can speak of doubling the temperature. From this example we learn that in 7.1.10 the effective Planck constant can be viewed both as setting the scale of length as compared to energy/momentum, or equivalently as setting the scale of energy/momentum as compared to space, and time. The relation 7.1.9 tells us therefore that, for more complex geometries, the same amount of electrons with space delocalization ξ will be obtained at a higher critical temperature, according to:

$$\frac{T_c(i)}{T_c(j)} = \frac{h_{\text{eff}}(i)}{h_{\text{eff}}(j)} . \tag{7.1.11}$$

In our theoretical framework, high critical temperatures show up as the consequence of the fact that, as expressed in 7.1.8, in superconductors with more complex geometrical structure, wavefunctions have a larger quantum uncertainty. In particular, keeping fixed all the other parameters, they have a larger $\langle \xi \rangle$. Therefore, the condition $n_S/n \geq (n_S/n)_0$ is satisfied at higher temperature.

We stress that the considerations about the introduction of an effective, geometry-dependent Planck constant concern the delocalization of wave functions. Namely, the role the Planck constant plays in the uncertainty relations, $\Delta x \Delta p \geq h/4\pi$, $\Delta t \Delta E \geq h/4\pi$, not the value of this constant as a conversion unit between energy and time, or space and momentum, in contexts not related to the uncertainty relations [3]. For instance, the energy levels, as computed through the Schrödinger equation, or a set of Schrödinger equations, out of a classical description of effective potentials, are computed using the ground value of the Planck constant. On the other hand, once the energy eigenvalues of a system are known, a geometry-dependent Planck constant must be used, in order to obtain the effective spreading of wavefunctions in a geometrically complex quantum system. To this regard, a consideration about the size of characteristic lengths which are introduced in the physics of superconductors, such as the coherence lengths ξ_0, $\xi(T)$, and the London penetration length λ, is in order. One could have the impression that, as we are keeping fixed the critical delocalization of wavefunctions at the transition to the superconducting phase, the entire classification about what are type I and what type II superconductors, discriminated by the ratio λ/ξ_0, has to be reconsidered. Indeed, this is not true: in this scenario all the classical results to this regard go through unchanged, because λ contains in its definition the Planck constant. In other words, both lengths λ and ξ_0 scale in the same way, and, as long as it is a matter of working with effective descriptions of superconductivity, such as for instance the Ginzburg-Landau effective theory, one can safely ignore rescalings, together with the grounds of a rescaling of the critical temperature.

[3] In other words, we could introduce a function of the geometry, which is set to one for flat geometry (or, to better say, for the ground geometry of the universe, corresponding to a curvature R of the order of the cosmological constant Λ):

$$\Delta x \Delta p \; [\Delta t \Delta E] \; \geq \; \frac{h}{4\pi} f(\langle R \rangle) \, , \qquad f(\langle R \rangle = \Lambda) \; = \; 1 \, . \tag{7.1.12}$$

In this way, the Planck constant remains formally invariant, while the ratios of above are expressed as ratios of different values of the function f.

7.2 Critical temperatures in various superconductors

In order to see up to what extent an approach based on our effective quantum gravity scenario can be applied, we consider now various examples of superconductors. The considerations of the previous section give us a clue on the role played by the geometry of a superconductor in determining its critical temperature. However, the detection of a regime of superconductivity is in general not a direct observation in itself: this regime is stated after the observation of several properties, such as for instance the magnetic properties. Magnetic effects play a relevant role also in the generation of an effective resistivity. Therefore, superconducting regime, and critical temperature in particular, may be very sensitive to effects such as impurities, and in general doping effects aimed to pin magnetic vortices. Also external conditions such as pressure do play in general a significant role. However, although strictly speaking also these conditions affect the geometry of the physical system and therefore act also at the quantum gravity level, in general these effects are of second order as compared to the changes they produce in the dynamic magnetic properties or other similar properties. Our investigation is therefore affected by a large amount of imprecision, and must be taken more as the indication of a tendency, than as a real precision test.

Low temperature superconductors are metals without a well defined structure. As mentioned in the introduction, in this case the order of magnitude of the critical temperatures is well predicted by BCS theory. Our concern will be with the "structured" configurations, characterized by higher critical temperatures. According to our previous discussion, in our approach we do not obtain an absolute determination of critical temperatures, but only of their ratios, as a function of the ratios of geometries. For our analysis, we will therefore use the low BCS temperatures as a reference point. We will start with mercury, which has a critical temperature around 4.2 K. The reason why we consider this element instead of other ones is that it allows a simpler derivation of the ratio of weights, in the sense of 7.1.4, to the next material we want to consider, NbTi.

7.2.0.4 Hg → NbTi

As a first test of the idea let us consider NbTi, the first step above Hg in the list of the table of page 290. In first approximation, the structure of NbTi should correspond to a Z_2 breaking of the symmetry of Hg. This would be exactly true if Nb and Ti had the same mass and properties. Indeed, we can ideally consider the symmetry breaking as roughly occurring through the pattern:

$$^{80}\mathrm{Hg} \xrightarrow{\sim Z_3} 2\,^{41}\mathrm{Nb} \xrightarrow{Z_1 \times Z_2} {}^{41}\mathrm{Nb} + {}^{22}\mathrm{Ti} + \cancel{^{22}\mathrm{Ti}}\,, \tag{7.2.1}$$

where Z_1 acts, as identity, on the first ^{41}Nb. This is somehow in between Z_2 and Z_3: it has less symmetry than a Z_2, but more than a Z_3, in that Nb looks like twice Ti, so that it ideally comes from the recombination of a Z_2 symmetry subgroup out of a breaking into three Ti. The critical temperature of NbTi should therefore lie somehow between 2 and 3 times the one of Hg: $T_c(\mathrm{NbTi}) \sim (8.4 + 12.6)/2 \sim 10.5\,\mathrm{K}$. Indeed, the observed critical temperature lies around 10 K. Of course, our evaluation has to be taken only as a rough, indicative estimate.

In passing from Hg to NbTi we have introduced a "weighted" breaking of the Hg molecular symmetry. The weight is precisely the mass of the atoms into which the initial homogeneous energy distribution breaks. This is justified by the fact that, in our theoretical framework, the geometries ψ in 2.1.16 are distributions of energy along space. The size of the mass of a particle depends on the weight the geometry (or the set of geometries) in which this particle appears has in the phase space of all the geometries. In turn, the weight of a geometry depends on the symmetry of the energy distribution. Approximately, the latter is "measured" by the space gradient of energy. Roughly speaking the higher is the density of energy gradient, the less homogeneous (= less symmetric) is the energy distribution. This can be understood as follows: let us consider a geometry, i.e. a particular distribution of energy along space, in its fundamental definition, as given in chapter 2, namely, as a map from a discrete space to a discrete space. At any time we move one energy unit (unit energy cell in the language of chapter 2) from a position in the target space to a neighbouring one, we modify one symmetry group factor. If we move just one cell we increase (or decrease) the energy gradient by two units, and break (restore) one "elementary" group factor. If we move another unit, we increase (decrease) further the energy gradient by two units, and act once again on another elementary group factor, and so on. The amount of increase/decrease of the energy gradient is therefore proportional to the factor of increase/reduction of the symmetry group of the configuration. These considerations are true for the configurations ψ entering in 2.1.16. However, owing to the properties of factorization of the phase space, and assuming that such a factorization is a good approximation when we want to "isolate" a local experiment such as those we are considering, we can transfer these global considerations also to the local description of superconductors. This implies neglecting the "extremely peripheral" geometries, anyway contributing for a minor correction, negligible for our present purposes. Therefore, instead of working with geometries as in 2.1.16, we work with "averaged" geometries as in 7.1.5, considering that everything outside the portion of universe we are testing remains unchanged. If we view geometries through an isomorphic representation in terms of symmetry groups:

$$\psi \leftrightarrow \prod_j G_j^{\psi} \,, \tag{7.2.2}$$

passing through the decomposition into an external and local part of the group:

$$\prod_j G_j^{\psi} \; = \; \left(\prod_j G_j^{\psi \, (ext)} \right) \times \prod_j G_j^{\psi \, (local)} \,, \tag{7.2.3}$$

it becomes clear that each geometry ψ_α can be factorized as:

$$\psi \; = \; \psi^{(ext)} \times \psi^{(local)} \,, \tag{7.2.4}$$

where "*local*" and "*ext*" precisely mean respectively the part of the geometry (or the corresponding symmetry group) describing the experiment (superconductor and related environment), and the rest of the universe. We can translate these considerations in terms of weights. Through the association:

$$\langle \psi^{(sc)} \rangle \longleftrightarrow \langle W^{(sc)} \rangle = \int_{\psi \in \{\psi^{(sc)}\}} \mathcal{D}\psi \, \mathrm{e}^{S} \left(\prod_i G_i^{\psi\,(ext)} \prod_j G_j^{\psi\,(local)} \right) , \tag{7.2.5}$$

(the label "(sc)" indicates the superconductor under consideration) we can use the factorization 7.2.4 to first integrate over the external part of every geometry. As long as the portion of universe represented by our experiment is small as compared to the rest of the universe, and isolated, in the sense that we can neglect the interaction of the system with the rest of the universe, the local and the external part of the geometries can be approximately treated as independent. Under this approximation, also the measure of integration can be factorized:

$$\mathcal{D}\psi \longrightarrow \mathcal{D}\psi^{(ext)} \times \mathcal{D}\psi^{(local)} . \tag{7.2.6}$$

We can therefore write:

$$\frac{1}{\mathcal{Z}} \int_{\psi \in \{\psi^{(sc)}\}} \mathcal{D}\psi \, \mathrm{e}^{S} \left(\prod_i G_i^{\psi\,(ext)} \prod_j G_j^{\psi\,(local)} \right) \approx \left\langle G^{(ext)} \right\rangle \times \langle G^{(local)} \rangle , \tag{7.2.7}$$

which allows to associate to $\langle \psi^{(sc)} \rangle$ a decomposition of weights:

$$\langle \psi^{(sc)} \rangle \to \langle W^{(sc)} \rangle \approx \langle W^{(ext)} \rangle \times \langle W^{(local)} \rangle . \tag{7.2.8}$$

This decomposition allows us to reduce the analysis of symmetries of geometries to just the crystal structure of our superconductors. The more, since superconductivity occurs as a property related to a characteristic length ξ, our considerations can be restricted to a region of this extension. In general, it is enough to look at a scale of order of the lattice length: the energy levels of the electrons are given in terms of collective wave functions [4], and all quasi-particle energies are measured in terms of the lattice length a, which sets

[4]For a review of these topics see for instance [109].

therefore the effective length/energy scale. In particular, when the energy gradient between neighbouring lattice periods is sufficiently smooth, it is possible to restrict the analysis to one lattice period. This is the case of the majority of the examples we are going to consider. With a certain degree of approximation, we can therefore write:

$$\langle W^{(local)} \rangle \;\; \propto \; \approx \int_a |\nabla E_i|_a \, . \tag{7.2.9}$$

This expression must be compared with the correction to the electron mass that in chapter 4 we referred to as quantum gravity corrections. As discussed there, the term ∇E has the form of a quantum mechanical correction in which energy depends on geometry as according to the Einstein's equations. 7.2.8 allows us to write then:

$$\frac{\langle W^{(sc)} \rangle_j}{\langle W^{(sc)} \rangle_i} \approx \frac{\int_{a_i} |\nabla E_i|_{a_i}}{\int_{a_j} |\nabla E_j|_{a_j}} \approx \frac{h_{\mathrm{eff}}(i)}{h_{\mathrm{eff}}(j)} \approx \frac{T_c(i)}{T_c(j)} \, . \tag{7.2.10}$$

Since we are talking of elements basically at rest, we can consider that the major contribution to the energy, determining the geometry of a configuration, comes from the rest energy, i.e. the mass. Therefore, to make the computation easier, instead of the integral of energy gradient we can consider the sum of the gradients of the mass distribution:

$$\int_a |\nabla_x E|_a \; \approx \; \sum_k^{(a)} |\Delta m^{(k)}| \, . \tag{7.2.11}$$

The ratios of critical temperatures between two such materials should then approximately be:

$$\frac{T_c(i)}{T_c(j)} \; \sim \; \frac{\sum_k^{(a_i)} |\Delta m_i^{(k)}|}{\sum_\ell^{(a_j)} |\Delta m_i^{(\ell)}|} \, . \tag{7.2.12}$$

This expression will allow us to investigate complex lattice structures. We stress that what matters is not simply the geometric lattice structure, with geometry intended as the space arrangement of atoms seen as massless geometric solids, but the space distribution of energies, in the sense of general relativity. If in first approximation we neglect isotope effects, in the purpose of comparing ratios of gradients, instead of the mass we may just consider the atomic number Z. We will now apply these considerations to the investigation of the next step in the table of page 290, Nb_3Sn.

7.2.0.5 Nb_3Sn

In the case of NbTi, the atomic numbers Z(Nb) = 41 and Z(Ti) = 81 lead to:

$$\sum |\Delta m| = |41 - 81| = 40 \qquad (\mathrm{NbTi})\,; \tag{7.2.13}$$

for Nb_3Sn, Z(Nb) = 41 and Z(Sn) = 50 give:

$$\sum |\Delta m| = |41 \times 3 - 50| = 73 \qquad (\mathrm{Nb_3Sn})\,; \tag{7.2.14}$$

The ratio of the sums of mass gradients of Nb_3Sn to NbTi is therefore 1.825, that, from 7.2.12 and $T_c(\mathrm{NbTi}) \sim 10\,\mathrm{K}$ should lead to some 18-19 K for the Nb_3Sn critical temperature. The observed one is around 18 K.

7.2.1 High-temperature superconductors

Although more complex, high-temperature superconductors are structured in layers, with a lattice structure that basically develops only along one coordinate. Their analysis is therefore, in first approximation, relatively simple, at least as long as one neglects the doping of certain sites with other elements. This introduces a further symmetry breaking that, in principle, leads to an enhancement of the estimated critical temperature. This operation may be considered somehow as a "built-in" ground effect, which underlies the properties of any one of these materials, and as such provides a systematic error, that can be observed in the general underestimate of the critical temperature. However, as doping varies from material to material, this further symmetry breaking cannot simply be "subtracted out" as a constant, universal effect: it introduces a further factor of uncertainty and approximation in our calculations. Our results should therefore be taken more for their capability to catch the main behaviour, than as an attempt to really provide a fine evaluation of the exact critical temperature. As a matter of fact, our estimates fall anyway within an error of at most 15% from the experimental observations.

7.2.1.1 LaOFeAs and SmOFeAs

For the group of iron-based superconductors we consider LaOFeAs and SmOFeAs. The crystal structure of LaOFeAs is arranged as a stack of layers in sequence (As) (Fe) (As) (La) (O) (La) etc. The one of SmOFeAs as a sequence of (As) (Fe) (As) (Sm) (O) (Sm) (see [110], [111], and [112]). The atomic numbers Z(La) = 57, Z(O) = 8, Z(Fe) = 26, Z(As) = 33 lead to:

$$\begin{aligned}\sum|\Delta m| &= 2|m(\mathrm{As})-m(\mathrm{Fe})| \\ &\qquad +|m(\mathrm{La})-m(\mathrm{As})|+2|m(\mathrm{La})-m(\mathrm{O})| \\ &\quad +|m(\mathrm{La})-m(\mathrm{As})| \\ &= 2|33-26|+|57-33|+2|57-8|+|57-33| \\ &= 14+24+98+24=160 \\ &\qquad (\mathrm{LaOFeAs})\,; \end{aligned} \tag{7.2.15}$$

Z(Sm) = 62, Z(O) = 8, Z(Fe) = 26, Z(As) = 33 lead to:

$$\begin{aligned}\sum|\Delta m| &= 2|m(\mathrm{As})-m(\mathrm{Fe})| \\ &\qquad +|m(\mathrm{Sm})-m(\mathrm{As})|+2|m(\mathrm{Sm})-m(\mathrm{O})| \\ &\quad +|m(\mathrm{Sm})-m(\mathrm{As})| \\ &= 2|33-26|+|62-33|+2|62-8|+|62-33| \\ &= 14+29+108+29=180 \\ &\qquad (\mathrm{SmOFeAs})\,; \end{aligned} \tag{7.2.16}$$

This gives as critical temperatures 42 K and 47 K respectively. The observed ones are 44 K and 57 K.

7.2.1.2 YBCO

We consider now the yttrium barium calcium copper oxide (YBCO) [113]. This material superconducts in its orthorhombic form. It is arranged as a stack of layers in sequence (Cu-O) (Ba-O) (Cu-O) (Y) (Cu-O) (Ba-O) (Cu-O). Differently from the previous examples, an evaluation of the mass gradients must here take into account also the fact that not only we have a gradient in passing from one layer to the neighbouring one, but also within each of the layers consisting of bonds of Ba and Cu with oxygen. In the planes presenting these bonds, it is not enough to just consider the gradient with the following plane: we must sum up also the mass gradient of the oxygen bond. On the other hand, in order to evaluate the overall gradient to be used in 7.2.12, it is not correct to sum up the absolute values of the "vertical" and the "horizontal" gradient. What counts for our purposes is the mean gradient contributed by each plane. We assume that, as in any propagation of errors, gradients in the two orthogonal axes sum up quadratically, as lengths of orthogonal vectors in a vector lattice. The overall gradient should approximately be given by the sum of the square roots of the quadratically propagated gradients of each layer, both in the "horizontal" and "vertical" directions. The evaluation of the mass gradient is complicated by the fact that, at the transition to the yttrium layer, oxygen couples both to copper and to yttrium, in an orthorhombic form. The crystal is therefore not structured in simple layers. In order to evaluate the mass gradient for the CuO_2-Y planes we make the approximation of attributing one oxygen atom to the copper layer, and one to yttrium. The expression of the sum of the mass gradients is then [5]:

$$\begin{aligned}\sum|\Delta m| &= 2\times\Big\{\sqrt{[(Cu+O)-(Ba+O)]^2+(Cu-O)^2}\\ &\quad+\sqrt{[(Ba+O)-(Cu+O)^2]+(Ba-O)^2}\\ &\quad+\sqrt{[(Cu+O)-O]^2+(Cu-O)^2}\\ &\quad+\ \sqrt{(O-Y)^2}\Big\}\,. \qquad (7.2.17)\end{aligned}$$

Considering the atomic numbers Z(Y) = 39, Z(Ba) = 56, Z(Cu) = 29, Z(O) = 8, we have $Cu + O = 37$, $Ba + O = 64$, and $Cu - O = 21$, $Ba - O = 48$, and therefore:

$$\begin{aligned}\sum|\Delta m| &= 2\times\Big\{\sqrt{(37-64)^2+21^2}+\sqrt{(64-37)^2+48^2}\\ &\qquad+\sqrt{(37-8)^2+21^2}\ +\ \sqrt{(8-39)^2}\Big\}\\ &\approx\ 2\times\{34.2+55.1+36+31\}\ \approx\ 312\\ &\qquad(\mathrm{YBCO})\,. \qquad (7.2.18)\end{aligned}$$

Rescaling the temperature from the previous elements through 7.2.12, we obtain a critical temperature $T_c \approx 312/160 \times 42 \sim 82\,\mathrm{K}$. If, in order to reduce the propagated error, instead of starting with the critical temperature of LaOFeAs as obtained through the series of rescalings from the metallic superconductors, we use as starting point its experimental value, 44 K, we obtain for YBCO a critical temperature of $\sim 86\,\mathrm{K}$. The experimental one is around 90-92 K.

The YBCO compound is part of a series, the so-called "123" superconductors, of similar critical temperatures, which differ by the substitution of yttrium with another element of the family of lanthanoids, including lanthanium. All these elements are heavier than yttrium, and we expect to find higher critical temperatures. This however is not always what happens. For instance, $(Y_{0.5}Gd_{0.5})Ba_2Cu_3O_7$ with T_c = 97 K, $(Y_{0.5}Tm_{0.5})Ba_2Cu_3O_7$ with 105 K, and $(Y_{0.5}Lu_{0.5})Ba_2Cu_3O_7$ with 107 K present an increasing critical temperature, as expected

[5]From now on we adopt the convention of indicating elements with Roman capital letters, and in italics their mass, so that e.g. Cu stays for m(Cu).

from the increasing of mass of the elements that substitute the pure yttrium, and the further symmetry breaking due to the fact that yttrium is substituted by a mixture of elements, as indicated in the brackets. However, $YbBa_2Cu_3O_7$ has $T_c = 89\,\mathrm{K}$, and $TmBa_2Cu_3O_7$ has $T_c = 90\,\mathrm{K}$ although Tm is lighter than Yb, and similarly $GdBa_2Cu_3O_7$ has $T_c = 94\,\mathrm{K}$, and $NdBa_2Cu_3O_7$ has $T_c = 96\,\mathrm{K}$, although Nd is lighter than Gd. A reason for this apparently odd behaviour could lie in the fact that the differences in atomic number are indeed very small, to the point that other effects play a non negligible role. In this case, in order to obtain more reliable predictions we would need a finer determination of the space layout of the energies and masses of these configurations.

There is another superconductor very similar to those of the YBCO series. It is $YSr_2Cu_3O_7$, which has a critical temperature $T_c = 62\,\mathrm{K}$. Strontium has atomic number 38, instead of the 56 of barium. In expression 7.2.18 we must therefore substitute $Ba + O = 64$ with 38 + 8 = 46, and $Ba + O = 48$ with $Sr - O = 30$. We have:

$$\begin{aligned} \sum |\Delta m| &= 2 \times \Big\{ \sqrt{(37-46)^2 + 21^2} + \sqrt{(46-37)^2 + 30^2} \\ &\qquad + \sqrt{(37-8)^2 + 21^2} + \sqrt{(8-39)^2} \Big\} \\ &\approx 2 \times \{22.9 + 31.3 + 36 + 31\} \approx 242.4 \\ &\qquad (\mathrm{YSrCCO})\,. \end{aligned} \tag{7.2.19}$$

Rescaling from YBCO, we obtain a critical temperature of 63-64 K, in substantial agreement with the experiments.

7.2.1.3 BSCCO

We consider now the bismuth-strontium-calcium-copper-oxide superconductors (BSCCO) [114]: Bi2212 ($Bi_2Sr_2CaCu_2O_2$) and Bi2223 ($Bi_2Sr_2Ca_2Cu_3O_{10}$). The lattice structure of the Bi2212 form is a stack of the following layers: (Bi-O) (Sr-O) (Cu-O_2) (Ca) (Cu-O_2) (Sr-O) (Bi-O) (Bi-O) (Sr-O) (Cu-O_2) (Ca) (Cu-O_2) (Sr-O) (Bi-O). The Bi2223 is similar, with one more (Ca) (Cu-O_2) layer. As for

YBCO, here too we must propagate both the "horizontal" and the "vertical" gradients. In this case the horizontal bonds are those of Bi, Sr and Cu with oxygen. For the Bi2212 form, we need therefore $Bi + O = 83 + 8 = 91$, $Sr + O = 38 + 8 = 46$, $Ca = 20$, $Cu + O_2 = 29 + 16 = 45$ and $Bi - O = 75$, $Sr - O = 30$, $Cu - O = 21$, to give:

$$\begin{aligned}\sum |\Delta m| &= 2 \times \Big\{ \sqrt{[(Bi+O)-(Sr+O)]^2 + (Bi-O)^2} \\ &\quad + \sqrt{[(Cu+2O)-(Sr+O)]^2 + (Sr-O)^2} \\ &\quad + \sqrt{[(Ca)-(Cu+2O)]^2 + [(Cu-O)+(Cu-O)]^2} \\ &= 2 \times \Big\{ \sqrt{(91-46)^2+75^2} + \sqrt{(45-46)^2+30^2} \\ &\qquad + \sqrt{(20-45)^2+(21+21)^2} \Big\} \\ &\approx 2 \times \{87.5+30+49\} \approx 332 \\ &\qquad (\mathrm{Bi2212})\,. \end{aligned} \tag{7.2.20}$$

Rescaling the temperature from LaOFeAs through 7.2.12 we obtain a critical temperature $T_c \approx 332/160 \times 42 \sim 87\,\mathrm{K}$. Starting from the experimental value, 44 K, in order to reduce the propagated error, we obtain for the Bi2212 a critical temperature of $\sim 91\,\mathrm{K}$, closer to the experimental one (92 K).

The structure of Bi2223 is very similar to the one of Bi2212, with just the difference of a Ca, CuO_2 layer-pair in each half-lattice block. In order to obtain the mass gradient of Bi2223 we must therefore just correct the former evaluation by adding an amount $|Ca - CuO_2)| + \sqrt{[Ca - CuO_2)]^2 + [2|Cu - O]^2}$:

$$\begin{aligned}\sum |\Delta m| &= \sum |\Delta m|(\mathrm{Bi(2212)}) + |20-45| \\ &\qquad + \sqrt{(20-45)^2+(21+21)^2} \\ &= 332 + 74 = 406 \\ &\qquad (\mathrm{Bi2223})\,, \end{aligned} \tag{7.2.21}$$

corresponding to a temperature of $406/160 \times 42 = 107\,\mathrm{K}$ ($\sim 111\,\mathrm{K}$ if we start from the experimental 44 K for the critical temperature of LaOFeAs). The experimental value is around 110 K.

7.2.2 The Tl-Ba-Ca-Cu-O superconductor

7.2.2.1 $Tl_2Ba_2CuO_6$ (Tl-2201)

The stacking sequence is as follows: (Tl-O) (Ba-O) (Cu-O_2) (Ba-O) (Tl-O) [6], and the expression of the mass gradient sum is:

$$\begin{aligned}
\sum|\Delta m| &= \sqrt{(Tl-O)^2+[(Tl+O)-(Ba+O)]^2} \\
&+ \sqrt{[(Ba+O)-(Cu+O+O)]^2+(Ba-O)^2} \\
&+ \big\{(Cu-O)^2+(Cu-O)^2 \\
&\qquad +[(Cu+O+O)-(Ba+O)]^2\big\}^{\frac{1}{2}} \\
&+ \sqrt{[(Ba+O)-(Tl+O)]^2+(Ba-O)^2} \\
&+ \sqrt{(Tl-O)^2+[(Tl+O)-(Tl+O)]^2}\,. \qquad (7.2.22)
\end{aligned}$$

From the atomic numbers Z(Tl) = 81, Z(Ba) = 56, Z(Cu) = 29 and Z(O) = 8 we derive $(Tl-O)=73$, $(Tl+O)=89$, $(Ba-O)=48$, $(Ba+O)=64$, $(Cu-O-O)=21$, and $(Cu+O+O)=45$. Plugging these values into the gradient sum expression, we obtain:

$$\begin{aligned}
\sum|\Delta m| &= \sqrt{73^2+(89-64)^2} \\
&+ \sqrt{(64-45)^2+48^2} \\
&+ \sqrt{2\times 21^2+(45-64)^2} \\
&+ \sqrt{(64-89)^2+48^2}\;+\;73 \\
&= 291\,. \qquad (7.2.23)
\end{aligned}$$

Rescaling now from LaOFeAs, expression 7.2.15, and using once again the 44 K of the experimental temperature, we obtain $(291/160)\times 44 = 80\,\mathrm{K}$ (had we used our calculated 42 K for LaOFeAs, we would have

[6] See refs. [115], [116], and also [117].

obtained $\sim 76.5\,\mathrm{K}$). The experimental critical temperature is around $80\,\mathrm{K}$.

7.2.2.2 $Tl_2Ba_2CaCu_2O_8$ (Tl-2212)

In this crystal there are two Cu-O-O layers with a Ca layer in between, with stacking sequence (Tl-O) (Ba-O) (Cu-O_2) (Ca) (Cu-O_2) (Ba-O) (Tl-O). In order to obtain the mass-gradient sum we have just to add to the previous computation a module accounting for the extra Ca layer vertically sandwiched between the two extra Cu-O-O, of which we consider also the horizontal contribution to the gradient:

$$\begin{aligned}\sqrt{[(Cu+O+O)-Ca]^2+(Cu-O)^2+(Cu-O)^2}\\ +\sqrt{[(Cu+O+O)-Ca]^2}\,.\end{aligned} \tag{7.2.24}$$

Considering that the atomic number of Ca is 20, this means an amount:

$$\sqrt{25^2+2\times 21^2}\;+\;25\;=\;64\,. \tag{7.2.25}$$

This gives a sum $291+64=355$, leading to a critical temperature of around $98\,\mathrm{K}$. The experimental one is around $108\,\mathrm{K}$.

7.2.2.3 $Tl_2Ba_2Ca_2Cu_3O_{10}$ (Tl-2223)

In this crystal there are three CuO_2 layers enclosing one Ca layer between each of them. That means, one more [(CU-O-O) Ca] module as compared to $Tl_2Ba_2CaCu_2O_8$. We obtain therefore a value of mass gradient sum $355+64=419$, leading to a critical temperature of $115\,\mathrm{K}$. The experimental one is $125\,\mathrm{K}$.

Both in this and in the previous superconductor we obtain slightly underestimated values of critical temperature. On the other hand, the ratio of the two critical temperatures we obtain, namely, 115/98, is in better agreement with the ratio of the experimental values. Indeed, it gives a slight overestimate, which partially compensates the underestimate of the first temperature. From a qualitative point of view, these under/over-estimates can be understood as follows: when a Ca layer is added to the $Tl_2Ba_2CuO_6$ structure, the symmetry of the configuration of a stack of "(X-O)" layers gets further broken, because the Ca layer does not contain an oxygen bond. Not taking this into account leads to an underestimate of the increase in critical temperature. On the other hand, when a further identical layer is added, there is a partial restoration of symmetry, which implies a reduction in the increase of critical temperature, thereby our over-estimation. This effect becomes more relevant in more complicated configurations: in Tl-based superconductors, the value of T_c decreases after four CuO_2 layers in $TlBa_2Ca_{n-1}Cu_nO_{2n+3}$, and in the $Tl_2Ba_2Ca_{n-1}Cu_nO_{2n+4}$ compound it decreases after three CuO_2 layers [118].

7.2.3 Comparing within families

Superconductivity is detected through investigation of the magnetic properties of materials. In particular, for what concerns high-temperature superconductors, pinning of magnetic flux through impurities plays a significant role, not only in reducing the effective resistance, and therefore affecting the conditions for the detection of a regime recognizable as the one of superconductivity, but, in the light of our analysis, also because it decreases the symmetry of the geometry. Also pressure plays a relevant role, because high pressures correspond to more remote geometries, and are expected to lead to higher critical temperatures (a fact that corresponds to the experimental observation). It is therefore rather difficult to give a correct quantitative account of the superconducting properties and the critical temperatures of all superconducting materials, and impossible to do it only in terms of comparison of average mass gradients referring to a single material taken as a universal starting point. In several cases, the best we can do is comparing critical temperatures within "families" of materials, which are assumed to share common properties, so that the change in the lattice structure taken into account by our evaluation of mass gradients can be considered as the only relevant variable and effective term of comparison.

7.2.3.1 Hg-Ba-Ca-Cu-O superconductor

An example of this kind of difficulties is provided by the Hg-series (Hg-1201, Hg-1212, Hg-1223 [119]). In principle, it is analogous to the series in which mercury is substituted by thallium (the Tl-series: Tl-1201, Tl-1212, Tl-1223), but, while the critical temperature of Tl-1201 is lower than 10 K, the one of the analogous compound made with Hg (one position lower in the atomic number scale) is around 94 K. Both these numbers escape the predictions we can make with our simple mass-gradient arguments, applied using mercury as starting point. Indeed the Hg-1201 material is a critical example in which doping plays a crucial role, whose details are still controversial. As reported in [120], depending on the amount of doping, this cuprate can superconduct or not, with a range of critical temperatures spanning the whole spectrum from zero to the maximal value. The critical temperature has proven to be also very sensitive to pressure [121]. In this case, the best we can do is to compare critical temperatures assuming comparable doping/flux pinning conditions. Assuming that, for instance, the highest critical temperature within the Hg-1201, 1212, 1223 series are obtained with a similar amount of such "external" inputs, we can expect to be able to give a reasonably good estimate of the ratios of critical temperatures *within* the Hg series. An illustration of the crystal structure of $HgBa_2CuO_4$ (Hg-1201, $T_c = 94\,K$), $HgBa_2CaCu_2O_6$ (Hg-1212, $T_c = 128\,K$) and $HgBa_2Ca_2Cu_3O_8$ (Hg-1223, $T_c = 134\,K$) can be found in [117]. Computing the ratios of temperatures along the same line as in the previous examples, we obtain:

$$\begin{aligned}\frac{T_c(Hg-1223)}{T_c(Hg-1212)} &\sim 1.23\,,\\ \frac{T_c(Hg-1212)}{T_c(Hg-1201)} &\sim 1.3\,,\\ \frac{T_c(Hg-1223)}{T_c(Hg-1201)} &\sim 1.59\,,\end{aligned} \qquad (7.2.26)$$

to be compared with the ratios of the experimental ones, namely 1.05, 1.36, and 1.43 for the (1223/1201) ratio. They show a similar situation of underestimate for the ratio of the lower pair of temperatures, and overestimate for the ratio of the third to the second one, as in the case of the thallium compound discussed above. Taking this into account, the ratios we find are not far from the experimental ones (the absolute determination of the temperature fails in this case to give a correct prediction, in that it would tell that both the thallium and the mercury -1201, -1212, -1223 series should have the same critical temperatures). This suggests that, keeping fixed all other conditions, the argument based on the evaluation of symmetry properties of the mass/energy configurations makes sense, although in some cases it is too simplified, and not sufficient to determine the overall conditions producing the particular state of a material which is detected as a regime of superconductivity.

A comparison restricted to elements belonging to the same family is our way of proceeding also in the case of higher temperature superconductors. Indeed, when passing to higher-T_c superconductors, and therefore to higher complexity of the lattice structure, a thorough analysis of the details of any part of the lattice block becomes the more and more difficult. On the other hand, in first approximation a detailed knowledge of the full lattice structure is not even necessary. The materials we are going to consider can be grouped into "families", whose elements share part of the lattice structure, and differ by the structure of just one (or some) of the lattice blocks. In this way, it is possible to perform a partial analysis, by comparing the critical temperatures among the members of each family. As the whole structure becomes longer and longer, it becomes smaller the error we introduce in weighting the various blocks according to their average length, thereby neglecting the details of the single mass gradients within common blocks. The difference from one material to the neighbouring one within a family usually consists in the substitution of some atomic elements, or in the addition of further replicas of already present layers. The mass differences introduced by these changes will be dealt with as a "second order" perturbation:

$$\frac{T_c'}{T_c} = \frac{T_c + \delta T_c}{T_c} = 1 + \frac{\delta T_c}{T_c} \approx 1 + \delta |(\nabla M)|_{\text{extra block}} \Big/ \sum |\nabla M| \,. \tag{7.2.27}$$

7.2.3.2 The SnBaCaCuO to (TlBa)BaCaCuO family.

i) *From* **160 K** *($Sn_3Ba_4Ca_2Cu_7O_\nu$) to* **200 K** *($Sn_6Ba_4Ca_2Cu_{10}O_\nu$).* The lattice structure of $Sn_3Ba_4Ca_2Cu_7O_y$ consists of a stack of (Ca) (CuO_x) [(Ba) (CuO_y) (Ba)] (CuO_x) (Ca) (CuO_x) [(Ba) (Sn-O) (Cu) (Sn-O) (Cu) (Sn-O) (Ba)] (CuO_x) , where in the first square bracket we indicate the light part of the lattice, in the second the heavy part, and (CuO_x), (CuO_y) indicate copper oxide layers. Here and in the following we use this notation to indicate, in general, (CuO_3) and (CuO_2) layers respectively [7]. The lattice structure of $Sn_6Ba_4Ca_2Cu_{10}O_\nu$ is obtained by doubling the "heavy" part of the lattice of $Sn_3Ba_4Ca_2Cu_7O_\nu$: instead of a sequence (Sn-O) (Cu) (Sn-O) (Cu) (Sn-O) we have now (Sn-O) (Cu) (Sn-O) (Cu) (Sn-O) (Cu) (Sn-O) (Cu) (Sn-O) (Cu) (Sn-O). The structure of this superconductor corresponds therefore to that of $Sn_3Ba_4Ca_2Cu_7O_\nu$, with the duplication of an entire lattice block. For the evaluation of the critical temperature we assume that, owing to the high number of lattice elements/layers, in first approximation we can consider the geo-

[7] Illustrations of this structure and of those of the following materials can be found ref. [122].

metry of the blocks structure as prevailing over the fine-structure of energy gradients, which distinguishes between light and heavy part of the lattice. That means, in first instance we deal with the blocks as if all lattice layers were equal, something that in the average is not far from the truth, and implies an error that becomes smaller and smaller as we go on with an increasing length of the crystal structure. Since the "replica" of the lattice block we add to obtain this superconductor corresponds to around 1/4 of the whole structure, we expect some 25% of increase in T_c from the one above, corresponding to an increase from 160 to 200 K [8].

ii) **212 K**: *(Sn$_5$In)Ba$_4$Ca$_2$Cu$_{10}$O$_\nu$*

The lattice structure consists of a stack of the following layers: (Ca) (CuO_x) [(Ba) (CuO_2) (Ba)] (CuO_x) (Ca) (CuO_x) [(Ba) (Sn,In-O) (Cu) (Sn,In-O) (Cu) (Sn,In-O) (Cu) (Sn,In-O) (Cu) (Sn,In-O) (Cu) (Sn,In-O) (Ba)] (CuO_x). We expect a higher T_c than the in last crystal of point (i) ($Sn_6Ba_4Ca_2Cu_{10}O_\nu$, $T_c =$ 200 K), as a consequence of the lower symmetry, now broken by the substitution of a tin atom with indium. This corresponds to a breaking of more or less one out of $18-20$ lattice layers, i.e. a $\sim$5-6 % of the total. Since the mass difference between Sn and In is of much lower order, in first approximation the increase of the critical temperature should be mainly determined by the symmetry breaking among different lattice layers, and therefore be of order $\sim$5-6 %. This gives indeed some 210-212 K, as is observed.

The mass difference between Sn and In plays a role as a second order effect, that can be observed in the smaller variation of the critical temperature after a change of the (Sn_5In) structure into (Sn_4In_2). The (Sn_5In) compound should superconduct

[8] More precisely, since the change is made in the heavy part of the lattice, it corresponds to more than 1/4 of the structure. However, since the modification consists in adding a replica of one layer, the effect is softened by the fact that there is also a further symmetry among the two identical layers. 1/4 is therefore to be taken as a rough estimate of the order of magnitude of the effect.

at a higher temperature than (Sn_4In_2), where there is a partial reconstruction of a higher symmetry within indium planes. Experimentally, one observes 212 K for the first, and 208 K for the second. Also in this case, an evaluation, even approximate, is rather difficult, because the naive value of 5/4 one would suppose (20% increase in the temperature) must be "tempered" by the fact that Sn and In weight almost the same. Their relative mass difference is 1/50, and this would mean a symmetry breaking of about 2%, indeed corresponding to the order of change in the observed T_c.

iii) **218 K**: *(Sn_5In)$Ba_4Ca_2Cu_{11}O_\nu$*

The lattice structure is similar to the one of (Sn_5In)$Ba_4Ca_2Cu_{10}O_\nu$ but contains one extra Cu in the light part of the lattice: (Ca) (CuO_x) [(Ba) (Cu_2O_y) (Ba)] (CuO_x) (Ca) (CuO_x) [(Ba) (Sn,In-O) (Cu) (Sn,In-O) (Cu) (Sn,In-O) (Cu) (Sn,In-O) (Cu) (Sn,In-O) (Cu) (Sn,In-O) (Ba)] (CuO_x). Adding a copper atom breaks part of the symmetry, thereby increasing T_c. As this occurs in one of the some 22 layers of each lattice block, we would expect this to produce a correction of about $\sim 1/22 = 4.5\%$ of T_c. This would mean some 9 K. However, in this estimate we don't consider finer corrections obtained by taking into account mass gradients. In practice, the breaking of symmetry is softened by the fact that there is a partial restoration of symmetry due to the fact that we are adding one more atom in a layer made of atoms of the same element. Indeed, the correction which is experimentally observed seems to be around 6 K, indicating a slightly lower symmetry breaking than in our rough estimate.

iv) **233 K**: *$Tl_5Ba_4Ca_2Cu_{11}O_\nu$*

The lattice structure is given by the following stack: (Ca) (CuO_x) [(Ba) (Cu_2O_y) (Ba)] (CuO_x) (Ca) (CuO_x) [(Ba) (Tl-O) (Cu) (Tl-O) (Cu) (Tl-O) (Cu) (Tl-O) (Cu) (Tl-O) (Ba)] (CuO_x). The heavy part of the lattice is similar to the one of the previous cuprate, with the suppression of the indium layer, and the substitution of tin atoms with thallium. In order to compare critical temperatures, let us compute the mass gradient sums corresponding to this part of the lattice for both these materials. They correspond to the stacking sequences (Ba) (Sn-O) (CuO) (Sn-O) (Cu) (Sn-O) (Cu) (Sn-O) (Cu) (Sn-O) (Cu) (Sn-O) (Ba) and (Ba) (Tl-O) (Cu) (Tl-O) (Cu) (Tl-O) (Cu) (Tl-O) (Cu) (Tl-O) (Ba) respectively. In the Sn sequence, one of the six tin atoms is substituted by an indium atom. Since the atomic numbers are respectively Z(In) = 49 and Z(Sn) = 50, in first approximation we neglect the slight asymmetry introduced by the (5 Sn)/In alternation. The mass gradient sums are:

$$\begin{aligned}\sum|\Delta m| &= |Ba-(Tl+O)| \\ &\quad +4\times\Big\{\sqrt{(Tl-O)^2+[(Tl+O)-Cu]^2} \\ &\quad +\ |Cu-(Tl+O)|\} \\ &\quad +\sqrt{(Tl-O)^2+[(Tl+O)-Ba]^2}\,, \end{aligned} \qquad (7.2.28)$$

and, neglecting the difference between Sn and In:

$$\begin{aligned}\sum|\Delta m| &= |Ba-(Sn+O)| \\ &\quad +5\times\Big\{\sqrt{(Sn-O)^2+[(Sn+O)-Cu]^2} \\ &\quad +|Cu-(Sn+O)|\} \\ &\quad +\sqrt{(Sn-O)^2+[(Sn+O)-Ba]^2}\,. \end{aligned} \qquad (7.2.29)$$

Inserting the atomic numbers Z(Ba) = 56, Z(Tl) = 81, Z(Cu) = 29, Z(O) = 8 and Z(Sn) = 50 we obtain:

$$\begin{aligned}\sum|\Delta m|(\mathrm{Tl}_5) &= |56-89| \\ &\quad +4\Big\{\times\sqrt{73^2+(89-29)^2}\ +\ |29-89|\Big\} \\ &\quad +\sqrt{73^2+(89-56)^2} \\ &= 33+4\times\{94.5+60\}+80,1 \\ &\approx\ 731\,, \end{aligned} \qquad (7.2.30)$$

and

$$\begin{aligned}\sum|\Delta m|(\mathrm{Sn_5In}) &= |56-58| \\ &\quad +5\left\{\times\sqrt{42^2+(58-29)^2}\;+\;|29-58|\right\} \\ &\quad +\sqrt{42^2+(58-56)^2} \\ &= 2+5\times\{51+29\}+42 \\ &\approx 444\,. \end{aligned} \tag{7.2.31}$$

The ratio between the two sums is therefore:

$$\frac{\sum|\Delta m|(\mathrm{Tl_5})}{\sum|\Delta m|(\mathrm{Sn_5In})} \approx 1.65\,. \tag{7.2.32}$$

In order to derive the rescaling of the critical temperature, we must see how much these gradients weight in the overall determination of the symmetry of these crystal configurations. The heavy part of the lattice amounts to more or less one half of the entire structure. However, a large part of this sub-lattice has a symmetry of five–almost six layers respectively. In practice, if the gradient (Ba)–(Tl-O), or (Ba)–(Sn-O) occurs on two stairs out of some 20–22, the change from (Sn-O)–(Cu) to (Tl-O)–(Cu), while occurring along some 5 layers, does not contribute so much to the reduction of symmetry. Owing to the symmetry of this stack, we expect it to contribute only by a factor $\sim \frac{1}{5!}$. Within the order of approximation we are making in this evaluation, it can therefore be neglected. The only part that counts is therefore the ratio:

$$\frac{\sqrt{[Ba-(Sn+O)]^2+(Sn-O)^2}}{\sqrt{[Ba-(Tl+O)]^2+(Tl-O)^2}} \approx 1.9\,, \tag{7.2.33}$$

that corresponds to the change in two out of some 20 layers, giving therefore a factor:

$$\frac{\langle||G||\rangle_{\mathrm{Tl_5}}}{\langle||G||\rangle_{\mathrm{Sn_5In}}} \approx \frac{1.9+10}{11} \sim 1.082\,, \tag{7.2.34}$$

implying a jump in critical temperature from 218 K to 236 K.

v) **242 K**: *(Tl_4Ba)$Ba_4Ca_2Cu_{11}O_\nu$*

The lattice structure is a stack of the following layers: (Ca) (CuO_x) [(Ba) (Cu_2O_y) (Ba)] (CuO_x) (Ca) (CuO_x) [(Ba) (X_1-O) (Cu) (X_2-O) (Cu) (X_3-O) (Cu) (X_4-O) (Cu) (X_5-O) (Ba)] (Cu), where, for every column, X_i stays four times for Tl, and one time for Ba in always different position for every layer. Between the cuprate of above and this one there is the substitution of some atoms of thallium with barium, which breaks part of the symmetry of the heavy part of the lattice. In this case, owing to the alternating position of the barium substituting thallium, the breaking of the symmetry, no more negligible as it was in the case of the Sn/In asymmetry, occurs not only in the "vertical" but also in the "horizontal" direction. In the aim of estimating the amount of symmetry breaking, we can make the approximation of considering just the effect of neighbouring lattice sites. In this approximation, each oxygen is surrounded either by four thallium, or by three thallium and one barium atom. The first case occurs only on one stair, whereas the other case occurs in the four remnant stairs. Therefore, we can roughly say that of the initial five thallium layers, four get separated into 1 (barium) plus 3 (thallium). In each of these four the symmetry factor is therefore $\frac{2}{3}$ (the barium/thallium mass ratio) $\times$ $\frac{4}{3}$ (the amount of remnant symmetry group, i.e. the ratio of the four before the breaking to the three after the breaking). All in all this makes:

$$\frac{5}{1 + 4\left(\frac{2}{3} \times \frac{4}{3}\right)} \approx 1.09756 \,. \tag{7.2.35}$$

Made on around 1/3 of the whole lattice raw, this implies a jump in the critical temperature of a factor around $(2 + 1.09756)/3$, that is, from the former 233 K-234 K to some 241 K-242 K, corresponding to the temperatures reported in the table of page 290.

vi) **254 K**: *(Tl_4Ba)$Ba_2Ca_2Cu_7O_{13+}$*

The lattice structure consists of a stack of (Ca) (CuO_y) (Ca) (CuO_x) [(Ba) (X_1-O) (Cu) (X_2-O) (Cu) (X_3-O) (Cu) (X_4-O)

(Cu) (X_5-O) (Ba)] (CuO_x). As compared to structure of (v), here the light part of the lattice has been partly cut out. In this case, differently from what one could expect, shortening a piece of the crystal structure leads to an increase of critical temperature. This can be understood as follows. All the elements of this family of materials are characterized by the fact of having a lattice structure consisting of a heavy and a light part. When considered from the point of view of a scale larger than just one lattice period, the reduction of the part with lighter masses, although in itself leading to a lower overall mass gradient within the single lattice length, owing to the shorter light-lattice structure, on a scale of several lattice units it increases the average gradient. The average effect is therefore equivalent to an increase of the heavy part of the lattice, the one with higher mass gradients. These situations are illustrated in figure 7.1. We can give a rough estimate of the effect, by considering that the light part has masses which are around one-half of those of the heavy part, and the change in the structure, as compared to the longer lattice form (the one of $(Tl_4Ba)Ba_4Ca_2Cu_{11}O_\nu$), amounts to suppressing some 2-3 layers in this light part, out of a total of ~ 20 lattice planes. This is a change of around 1/2 of 10%, i.e. $\sim 5\%$, corresponding to a jump in the temperature of some 12 K. This leads from the former 242 K to around 254 K, the value experimentally observed (see table of page 290).

This example shows that, although working within a single unit of lattice length, as implied in 7.2.10–7.2.12, is in most cases correct, the comparison of geometries is in principle something more subtle. In the case of $(Tl_4Ba)Ba_2Ca_2Cu_7O_{13+}$, just considering one unit of lattice length is not enough.

7.2.3.3 *The (SnPbIn)BaTmCuO family: from 163 K to 195 K.*

The lattice structure of $(Sn_{1.0}Pb_{0.5}In_{0.5})Ba_4Tm_4Cu_6O_{18+}$, $T_c = 163\,K$, consists of a stack of (0.5($Sn_{1.0}Pb_{0.5}In_{0.5}$)-O) (Ba) (CuO_x) (Tm) (CuO_x) (Ba) (0.5($Sn_{1.0}Pb_{0.5}In_{0.5}$)-O) (Ba) (CuO_x) (Tm) (CuO_y) (Tm)

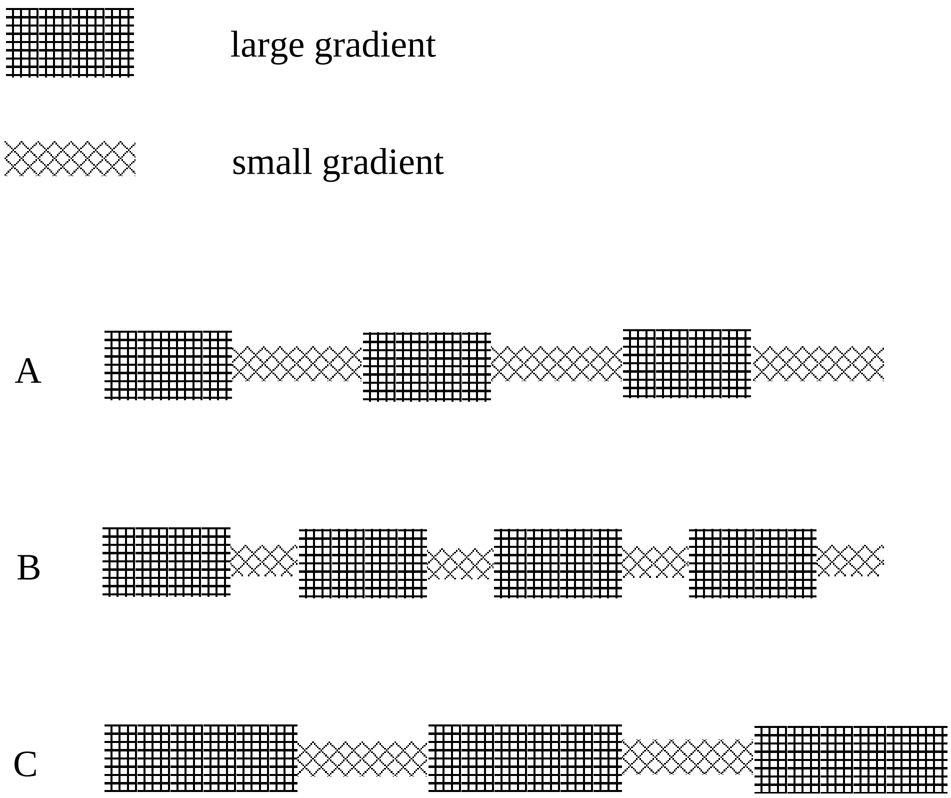


Figure 7.1: Both shortening the light, small mass gradient part (example B), and lengthening the heavy, high gradient part of the lattice (example C) lead to an effective increase of average mass gradient as compared to A, and therefore to a higher remoteness of the geometry, which reflects in an increased critical temperature.

(CuO_y) (Tm) (CuO_x) (Ba). The lattice structure of $(Sn_{1.0}Pb_{0.5}In_{0.5})$ $Ba_4Tm_5Cu_7O_{20+}$, $T_c = 185\,K$, consists of a stack of $(0.5(Sn_{1.0}Pb_{0.5}In_{0.5})$-O) (Ba) (CuO_x) (Tm) (CuO_x) (Ba) $(0.5(Sn_{1.0}Pb_{0.5}In_{0.5})$-O) (Ba) (CuO_x) (Tm) (CuO_y) (Tm) (CuO_y) (Tm) (CuO_y) (Tm) (CuO_x) (Ba).
The lattice structure of $(Sn_{1.0}Pb_{0.5}In_{0.5})Ba_4Tm_6Cu_8O_{22+}$, $T_c = 195\,K$, consists of a stack of $(0.5(Sn_{1.0}Pb_{0.5}In_{0.5})$-O) (Ba) (CuO_x) (Tm) (CuO_x) (Ba) $(0.5(Sn_{1.0}Pb_{0.5}In_{0.5})$-O) (Ba) (CuO_x) (Tm) (CuO_y) (Tm) (CuO_y) (Tm) (CuO_y) (Tm) (CuO_y) (Tm) (CuO_x) (Ba). It is difficult to compare with the elements of the previous series. On the other hand, since the differences among the elements of this series consist in adding a [(CuO_y) (Tm)] pair of layers within the same lattice subset, it is relatively easy to compare the elements within this group. In practice we are adding a Tm line at each step, increasing the lattice complexity by one layer out of a total of 10 in the first case, and of 10+1 in the second case. We expect therefore an increase of T_c by a factor 11/10 and 12/11 respectively. This corresponds to a jump from $\sim$163 K to $\sim$180 K, and to $\sim$195 K in the second case, to be compared with the values 163 K, 185 K and 195 K reported in the table of page 290.

7.3 To summarize

This analysis provides support to the hypothesis that quantum gravity effects may be at the ground of the understanding of the relation between lattice complexity and critical temperature of superconductors. Roughly speaking, working in a quantum gravity framework effectively means having a Planck constant dependent on the *gradient* of the distribution of energy along space. If we introduce an energy density $\rho(E)$, this in practice means that we are effectively promoting $\hbar$ to:

$$\hbar \;\to\; \hbar(\nabla\rho(E))\,. \tag{7.3.1}$$

Equivalently, we can also say that we work with a geometry-dependent Planck constant:

$$\hbar \;\to\; \hbar(g_{\mu\nu})\,, \tag{7.3.2}$$

or, to work with quantities independent on the choice of coordinate system, with a curvature-dependent Planck constant. Although all the expressions considered in this chapter are worked out within a non-field theoretical framework, from a heuristic point of view this dependence can be understood as follows. Quantization of gravity introduces an effective dependence on $\hbar$ in the modes of propagation of the metric tensor $g_{\mu\nu}$. This means that, even if we start with a space with a classical background metric, after quantization, and as a consequence of the back-reaction due to the interaction with matter and radiation, we will end up with a space with $\hbar$-dependent geometry, $g_{\mu\nu}(\hbar)$. Taking the point of view of considering geometry as a primary, independent input corresponds to inverting the relation $g_{\mu\nu} = g_{\mu\nu}(\hbar)$ to $\hbar = \hbar(g_{\mu\nu})$. The functional dependence is not simple; on the other hand, its explicit expression is not even fundamental, because it expresses only an effective parametrization: in general, in order to derive, case by case, the appropriate effective parametrization, one has to refer to 2.1.16. The ground value of the Planck constant is the one corresponding to the "vacuum", which in our case is the universe with uniform curvature, corresponding to the cosmological constant [9]. A uniform curvature gives a universal contribution that can be subtracted, i.e. re-absorbed into a redefinition of the Planck constant. This is what is done when gravity is decoupled from the quantum theory, and one recovers the traditional quantum theory.

A dependence of the Planck constant on the geometry means that also the amount of quantum delocalization of wave functions, the mechanism at the ground of superconductivity, depends on the geometry. However, the relation between critical temperature and lattice complexity of superconductors cannot be observed in a clean, direct way: superconductivity is a regime in most cases "unstable" in pure materials, and the way it is detected makes measurements very sensitive to several additional conditions. The relation we propose between criti-

[9] More precisely, the ground curvature is the average sum of the cosmological term, plus the contributions of matter and radiation. It is the sum of all these terms what gives the universe the ground average curvature of a 3-sphere (see chapter 5).

cal temperature and quantum geometry only works at the net of any other effect, such as degree of doping/pinning of magnetic flux, etc. A quantitative prediction is only possible when the contribution of these effects can been subtracted. The agreement between observations and our theoretical predictions has therefore to be read "in the average", and works better when comparing temperatures between materials belonging to the same "family", for which the other conditions can be assumed to be similar (the case of the Hg-1201/1212/1223 series is exemplar of this situation). Nevertheless, the agreement between predictions and experimental observations is impressive. Our analysis provides a further indication that, differently from what one is used to expect, quantum gravity is not just a matter of Planck scale phenomena, but in principle comes into play, to contribute for non-negligible corrections, in any quantum system corresponding to a non-trivial geometry of space-time.

Since in this scenario masses, couplings, and the geometry itself evolves with time, also the effective Planck constant expressed in 7.3.2 is expected to evolve with time, through the time dependence of the metric $g_{\mu\nu} \to g_{\mu\nu}(t)$. Time dependence of the metric is familiar in general relativity, where it occurs through its dependence on the space-time coordinates. However, here we mean something more subtle: we mean that also the microscopic structure of a crystal, and its energy gradients, evolve with time. In chapter 4 we have seen that masses evolve as negative powers of the age of the universe. As also seen in chapter 6, this implies that the ratio of different mass scales increases with time. We must therefore expect also an increase in the ratio of different degrees of delocalization of wavefunctions. How fast should this go can be estimated by considering that, approximately, mass ratios scale as powers of the age of the universe. With a similar degree of approximation, we can assume that also lattice gradient ratios scale as powers of the age of the universe. A factor 2 in the ratio of the mean weights of lattice geometries at present time:

$$\frac{\xi_i}{\xi_j} \approx \frac{\langle \nabla m_i \rangle}{\langle \nabla m_j \rangle} \sim 2 \,,$$

corresponds to a very small exponent $a_{(\xi_i/\xi_j)}$ of the evolution:

$$\frac{\xi_i}{\xi_j} \approx \mathcal{T}^{a_{(\xi_i/\xi_j)}} .$$

This is given in fact as $\log 2 = a_{(\xi_i/\xi_j)} \log \mathcal{T}$, where the age of the universe $\mathcal{T}$ is expressed in units of appropriately converted Planck length. At present time $\mathcal{T} \sim 10^{61}$. This kind of evolution is therefore only detectable on a large, cosmological, time scale, and negligible for usual purposes.

Transition Temperature in Kelvin	Material	Class
254	$(Tl_4\ Ba)\ Ba_2\ Ca_2\ Cu_7\ O_{13+}$	Copper-oxide superconductors
242	$(Tl_4\ Ba)\ Ba_4\ Ca_2\ Cu_{11}\ O_\nu$	
233	$Tl_5\ Ba_4\ Ca_2\ Cu_{11}\ O_\nu$	
218	$(Sn_5\ In)\ Ba_4\ Ca_2\ Cu_{11}\ O_\nu$	
212	$(Sn_5\ In)\ Ba_4\ Ca_2\ Cu_{10}\ O_\nu$	
200	$Sn_6\ Ba_4\ Ca_2\ Cu_{10}\ O_\nu$	
160	$Sn_3\ Ba_4\ Ca_2\ Cu_7\ O_\nu$	
195	$(Sn_{1.0}\ Pb_{0.5}\ In_{0.5})Ba_4Tm_6Cu_8O_{22+}$	
185	$(Sn_{1.0}\ Pb_{0.5}\ In_{0.5})Ba_4Tm_5Cu_7O_{20+}$	
163	$(Sn_{1.0}\ Pb_{0.5}\ In_{0.5})Ba_4Tm_4Cu_6O_{18+}$	
125	$Tl_2Ba_2Ca_2Cu_3O_{10}$	
108	$Tl_2Ba_2CaCu_2O_8$	
80	$Tl_2Ba_2CuO_6$	
110	$Bi_2\ Sr_2\ Ca_2\ Cu_3\ O_{10}$(Bi2223)	
92	$Bi_2Sr_2CaCu_2O_2$ (Bi2212)	
92	$YBa_2\ Cu_3\ O_7$ (YBCO)	
57	SmFeAs(O,F)=SmOFeAs	Iron-based superconductors
44	LaFeAs(O,F)=LaOFeAs	
18	$Nb_3\ Sn$	Metallic low-temp. superconductors
10	NbTi	
4.2	Hg	

8 Prime numbers and the structures of the universe

In order to recover the ordinary description of physics, in the previous chapters we have considered the limit to the continuum. Traditionally, the basic bricks of the description of the physical world are in fact the plane-wave free asymptotic states. Their interaction is dealt with as a perturbation. This approach proves to be successful in the description of weak forces (weak and electroweak), as well as in the case of "large-scale", classical gravitation (although excluding the cosmological scale of the evolution of the universe, in which case the small quantum gravity effects sum up on the long distance and large time elapse). Our approach however provides us with a non-perturbative description of the universe, in which the actual universe results from the superposition of geometries weighted according to their statistical weight in the phase space of all the geometries. This weight corresponds to the volume of their combinatorial group, i.e. to the number of ways they can be formed, and is in turn related to the frequency with which they therefore do occur. In this approach, a relevant concept is therefore that of factorization of the weight of a geometry into prime factors, that we interpret as corresponding to the weight of the elementary structures of this universe. These structures have in principle nothing to do with the traditional physical elementary structures, such as for instance the elementary particles. Therefore, when it is a matter of describing a free electron, it is still convenient to switch to its quantum mechanical description as a free wave. But there are cases in which this other kind of elementary structures is more appropriate. In particular, when a truly non-perturbative description is required, at least for the investigation of certain properties. We will see how this will allow us to get a new insight into the scaling structure of the universe, and to derive the scaling behaviour of the weak and strong coupling.

8.1 Prime numbers and complexity of structures

It is a common observation that, in first approximation, the universe seems to reproduce its shapes at different scales. For instance, a planet surrounded by its satellites looks like a kind of miniature-version of the solar system. Similarly, although in a very loose way, it is not completely wrong to imagine the atom as a small solar system. In first approximation, this appears to be due to the fact that also the electric force behaves in a similar way to the gravitational one, both at the classical level of the Coulomb-like expression of the potential, and at a field-theoretical level, being both photon and graviton massless fields. But here we want to understand *why* the physical world is ruled by forces that in first approximation behave in a similar way, reproducing similar structures at different scales.

In our scenario, the universe, and therefore any physical system, is given by the superposition of an infinite number of geometries, each one with a different weight. If we want to look at the scale properties in order to see whether and why certain structures and shapes are roughly reproduced at different scales, we must first of all consider the average over the staple of geometries, i.e. the "mean geometry", contributing to form the universe at a certain scale, and then also mod out by the structures at lower scales. This last operation is required by the fact that, when for instance we compare a planet and its satellites with the solar system, we neglect the fact that certain elements of the solar system, namely certain planets, have themselves in turn the structure of small solar systems, and so on.

We want to obtain the number of elementary structures around a time/energy scale N. At any energy scale N the most entropic geometry is the 3-sphere of radius N (see chapter 2). Its weight scales as $\exp N^2$ times a factor depending on the total volume of space, and a trivial factor $N!$, common to all the geometries at energy/time N.

In our discussion these factors, which account for a permutation of indistinguishable energy units, and for the number of possibilities of placing the center of the 3-sphere along a space of finite extension, are always implicitly factored out. The relevant term, $\exp N^2$, which is the part of the weight depending on the intrinsic symmetry group of the sphere, has to be intended as the appropriate natural integer whose size scales as the exponential of the square of the radius: although we use the expression $\exp N^2$, here we are indeed always speaking of an integer number. As discussed in chapter 2, in this setup one works always in a space regularized by a cut-off, to be eventually removed, which sets the volume of space and the number of dimensions to finite values. Under these conditions, as long as the cut-off sets a target volume much larger than the one of the sphere, the extra factor is almost the same for all the geometries with a volume close to the one of the 3-sphere, and can be factored-out. The contribution due to the cut-off becomes relevant for the geometries in which the units of energy are distributed along a very large volume, much larger than the one of the 3-sphere. On the other hand, as it has been discussed in chapter 2, the weight of these geometries is much lower than the one of the 3-sphere, which alone weights more than the sum of all the other geometries [1]. In our analysis, we can therefore *normalize* all the weights dividing by the extra-factor of the 3-sphere, so that the weight of the 3-sphere is simply $\exp N^2$. This will introduce non-integer weights, but since we are interested in the scaling properties, what counts here is the relative scaling of subsets of numbers, and this can be investigated independently on the normalization we choose. The error due to the cut-off can be made arbitrarily small. In the limit of infinite volume of the target space, the volume to be factored out becomes the same for all the geometries. To better say, once the overall volume factor is factored out the distance between the actual weight W and the closest integer number, $n(W)$, vanishes in this limit: $|W - n(W)| < \mathcal{O}(1/V)$, where V is the volume of the target space

[1] We can also safely restrict our considerations to three dimensions, because the weights of the spheres at different dimensionalities, which are anyway the most entropic geometries for each dimension, are exponentially suppressed and therefore contribute to corrections of much lower order.

(not to be confused with the volume of the 3-sphere, $\sim N^3$, which corresponds to the volume of the classical geometry of the universe), so that $|W - n(W)| \stackrel{V \to \infty}{\longrightarrow} 0$.

For what we have just discussed, at any physical energy scale N we can associate an integer n of approximate size $\sim \exp N^2$. Let us indicate with $\pi(n)$, as is usual, the number of primes up to the integer n. The quantity of interest for us is the number of primes around n: $d\pi(n(N))/dN$ $(\times \Delta N = 1)$. This precisely indicates the number of independent, basic structures, around the chosen scale, neglecting higher or lower scales. In order to simplify the computation, instead of the finite interval we consider the derivative, which gives us the increment in the number of structures per increment of the scale. Consider the approximate formula giving the number of primes up to the integer n:

$$\pi(n) \approx \frac{n}{\ln n} \,. \tag{8.1.1}$$

According to the theorem of primes, this approximate equation is the more and more exactly satisfied the larger and larger is the size of n. By inserting $n = \exp N^2$, and taking the derivative with respect to N, we obtain:

$$\frac{d\pi(n(N))}{dN} = \frac{d}{dN}\frac{\mathrm{e}^{N^2}}{N^2} = \frac{2N\mathrm{e}^{N^2}}{N^2} - 2\frac{\mathrm{e}^{N^2}}{N^3} \,. \tag{8.1.2}$$

In order to compare the behaviour at different scales we must then normalize the increments of our differential expression dividing by the scale N itself. We obtain:

$$\frac{d\pi(n(N))}{d\ln N} = \frac{1}{N}\frac{d\pi(n(N))}{dN} \approx \frac{2\mathrm{e}^{N^2}}{N^2} - 2\frac{\mathrm{e}^{N^2}}{N^4} \,. \tag{8.1.3}$$

We now mod out the number of structures at the lower scale, by dividing by $\pi(n(N))$, finally obtaining the expression we were looking for:

$$\frac{d\ln\pi(n(N))}{d\ln N} \approx 2 - \frac{2}{N^2} \,, \tag{8.1.4}$$

where we used the symbol $\approx$ in order to make clear that this is only an approximated expression, obtained by considering just the most entropic geometry. In first approximation, the r.h.s. of expression 8.1.4 is a constant. This tells us that, roughly, the world shows up with similar structures at different scales. Roughly speaking, one could say that, if one forgets quantum corrections (i.e. the contribution of the rest of the staple of geometries out of the classical one), "an atom is like a solar system", thereby justifying the Bohr planetary-like approximation of the atom. The second term in the r.h.s. of 8.1.4 comes from the logarithmic factor, which characterises the distribution of primes, singling them out of the whole set of natural numbers. It gives a $1/N^2$ correction, that looks negligible at large N. However, this correction is, depending on the scale, precisely of the order either of the quantum corrections, or of the corrections introduced in the classical geometry by matter clusters (observe also that the energy density of the universe scales like $1/N^2$). As we are going to discuss in the next section, this term can be considered an "interaction" term, that tells about the strength of medium and large range forces. The way it scales with N tells us that at larger scales the world tends to become more "simple" in the sense of more classical and flat.

8.2 The scaling of couplings

Knowing the distribution of prime vs non-prime numbers allows us to derive also certain scaling properties of the couplings. As discussed, in this theoretical framework a coupling is a volume in the phase space of the geometric configurations of the universe: it measures the weight of a transformation of particles. Along the evolution of the universe couplings scale therefore basically like ratios of masses. However, physics is more complex than just direct transitions from particle A to particle B. Indeed, we distinguish between long range and short range forces, and between strong and weak forces. The turning point between these two is the unit of measure of all the scales: the Planck scale. The gravitational coupling has here by definition size 1 (see chapter 3). If the strength of the gravitational coupling is fixed, the strength of the electroweak coupling has been derived in chapter 4 by going to a logarithmic representation of the physical world. As discussed in chapter 3, this is the picture in which gravity is decoupled, and one can easily investigate the spectrum of the elementary particles. Once obtained the bare value of the electroweak coupling from a ratio of volumes at a certain age of the universe, the actual value at the appropriate physical scale has then been computed by running the bare value from the ground scale of masses, assuming a logarithmic rescaling of the coupling as a function of the energy scale.

In the light of the present analysis, we can get a further insight in what we are precisely doing when passing to a perturbative representation. Within the set of all possible geometries, a special role is played by those which have a weight that, once normalized to the 3-sphere as above, is given by a prime number [2]. They don't contain subsets corresponding to subgroups of their global symmetry group. As such, they must be viewed as "global" geometries: they describe the entire universe as a whole piece. We can test this interpretation by considering that, as compared to the other geometries, the "local" ones, the volume of their symmetry group should loose a factor corresponding to the volume of the universe. The weights of the global geometries must therefore roughly scale as $1/N^3$ of the weights of the local configurations. The heaviest local geometry is the 3-sphere (the weight of the 3-sphere clearly cannot be a prime number, because the symmetry group of the sphere has subgroups, whose weight is an integer divisor of the weight of the sphere). As discussed in chapter 2, the weight of the 3-sphere is of the order of the entire sum of weights, that we indicate as $\mathcal{W}(N)$. If we indicate with $\mathcal{W}_{\text{global}}(N)$ the sum of the weights of the global geometries at time (or energy) N ([3]), we have that this scales approximately as:

$$\mathcal{W}_{\text{global}}(N) \approx \frac{\mathcal{W}(N)}{N^3} \approx \frac{\mathrm{e}^{N^2}}{N^3}\,, \tag{8.2.1}$$

[2]The fact that 2.1.16 sums by definition over all possible geometries ensures us that such geometries do exist "by construction".

[3]The total weight is also the total number of ways the N units of energy can be distributed along space.

where we have approximated $\mathcal{W}(N) \approx \mathrm{e}^{N^2}$. Integrating over time, this gives a scaling:

$$\int_N \mathcal{W}_{\mathrm{global}}(n) \approx \frac{\mathcal{W}(N)}{N^2} \approx \frac{\mathcal{W}(N)}{\ln \mathcal{W}(N)} \,. \tag{8.2.2}$$

This expresses the relation between the total weight, up to the size $\mathcal{W}(N)$, of the global geometries, and the total weight of all the geometries. With the substitutions $\pi(n) \leftrightarrow \int_N \mathcal{W}_{\mathrm{global}}$ and $n \leftrightarrow \mathcal{W}(N)$, this is the same as the relation between the number of primes and the natural numbers, expression 8.1.1. As previously discussed, as long as the regularization cut-off V is finite this is just a correspondence between the scaling behaviour of weights and sets of numbers. It becomes however an exact correspondence with the sets of natural and prime numbers in the limit in which the cut-off is removed by factorizing out V, i.e. the limit $V \to \infty$, when the weights become exactly integer numbers.

Decoupling gravity from the theory, and in particular separating the effects of gravity on the weak couplings, corresponds to looking only at the geometries that describe the long-range part of the interaction, related to the non-local geometries. Massive objects correspond instead to localizable objects, and clearly belong to the local part of the set of geometries [4]. The strength of the coupling is related to the weight of this subset of geometries. Looking at its running through the mass scales means considering the weight of this subset of geometries relative to the weight of the geometries building up the gravity part:

$$\alpha \leftrightarrow \frac{\int_M \mathcal{W}_{\mathrm{global}}(m)}{\mathcal{W}(M)} \approx \frac{1}{\ln \mathcal{W}(M)} \quad \Longrightarrow \alpha^{-1} \sim \ln \mu \,. \tag{8.2.3}$$

The actual energy scale μ is not the total energy of the universe, N: microscopic energy scales are a fraction of the total energy of the universe, produced by the fact that in the microscopic physics one looks

[4] Notice that the usual field-theoretical description of free states in terms of plane waves, which are by definition infinitely extended, indeed assumes this point of view.

just at a subregion of each geometric configuration. Rather than being the actual value of a coupling, expression 8.2.3 has to be understood as giving the scaling behaviour with respect to the energy scales. The logarithmic running of couplings catches the scaling of the long-range part of the interactions. It gives quite correctly the behaviour of the electroweak coupling through the energy scales *at fixed age of the universe* [5]. We get here therefore another way of understanding why in the perturbative theory masses are free parameters: to be rigorous, the perturbative theory doesn't know about masses, consistently with the fact that they cannot be directly introduced in perturbative field theory. The perturbative, logarithmic running belongs to a long-range, flat-space description of the world.

Let us now consider a strong force, like the colour interaction in the strong coupling regime (which is the dominant regime at sub-Planckian energies, see chapter 3). In this case, the interaction is not global, i.e. of infinite-range, but involves only localized objects. The strength of the coupling is therefore related to the part of numbers which are not prime, with density $\sim 1 - \frac{1}{\ln n}$. As a consequence, it does not have the logarithmic scaling of a perturbative, gravity-decoupled picture. It just evolves according to a power-low time-dependence on the age of the universe, $\sim n^{\beta}$ for some exponent β), like the different mass scales do (see chapter 4). A correct treatment of the strong force *and* the weak interaction can only occur within a theoretical framework in which also gravity is considered, i.e. a theory in which also localized massive objects are consistently described.

[5]The weak interaction is here a medium range interaction which consists of a "long range part", the pure coupling, which behaves, and scales, similarly to the electromagnetic coupling, and a suppressing mass term, which works as a kind of cut-off, so that the effective coupling is α_W / M_W^2. The scaling of α_W is logarithmic.

Appendix

Conversion units for the age of the universe

We give here some conversion factors from time units to Planck mass units. When expressed in seconds, one year is:

$$1 \text{ year (yr)} = 3.1536 \times 10^7 \text{ s}.$$

In order to convert this value to eV units we divide by $\hbar = 6.582122 \times 10^{-22}$ MeV s. We obtain:

$$1 \text{ yr} = 4.791160054 \times 10^{28} \text{ MeV}^{-1}.$$

Considering that the Planck mass $\mathrm{M_P} = 1.2 \times 10^{19}$ GeV, we have also the relation:

$$1 \text{ yr} = 3.992633379 \times 10^{50} \, \mathrm{M_P^{-1}}.$$

The age of the universe $\mathcal{T}$, estimated to be around 11.5 to 14 billion years, reads therefore:

$$\mathcal{T} \approx \left\{ \begin{array}{c} 4.59152839 \\ 5.58968673 \end{array} \right. \times 10^{60} \, \mathrm{M_P^{-1}}.$$

If instead we take the neutron mass as the most precise way of deriving the age of the universe, from expression 4.3.26 and the present-day measured neutron mass, we obtain:

$$\mathcal{T} \approx 5.038816199 \times 10^{60} \, \mathrm{M_P^{-1}} \quad (= 12.6202827 \times 10^9 \text{ yr}). \quad \text{(A.1)}$$

References

[1] A. Gregori, C. Kounnas, and P. M. Petropoulos, "Non-perturbative gravitational corrections in a class of n = 2 string duals," *Nucl. Phys.*, vol. B537, pp. 317–343, 1999.

[2] A. Gregori *et al.*, "R**2 corrections and non-perturbative dualities of n = 4 string ground states," *Nucl. Phys.*, vol. B510, pp. 423–476, 1998.

[3] A. Gregori, C. Kounnas, and P. M. Petropoulos, "Non-perturbative triality in heterotic and type ii n = 2 strings," *Nucl. Phys.*, vol. B553, pp. 108–132, 1999.

[4] A. Gregori, C. Kounnas, and J. Rizos, "Classification of the n = 2, z(2) x z(2)-symmetric type ii orbifolds and their type ii asymmetric duals," *Nucl. Phys.*, vol. B549, pp. 16–62, 1999.

[5] A. Gregori, "Rank 48 gauge group in heterotic string," 2000.

[6] A. Gregori, "String-string triality for d=4, z_2 orbifolds," *JHEP*, vol. 06, p. 041, 2002.

[7] A. Gregori, "Naturally time dependent cosmological constant in string theory," 2004.

[8] A. Gregori, "Entropy, string theory, and our world," *hep-th/0207195*, 2002.

[9] A. Gregori, "On the time dependence of fundamental constants," 2002.

[10] A. Gregori, "An entropy-weighted sum over non-perturbative vacua," *arXiv:0705.1130*, 2007.

[11] A. Gregori, "A note on the phases of natural evolution," 2007.

[12] A. Gregori, "Relativity as classical limit in a combinatorial sce-

nario," 2009.

[13] A. Gregori, "Is the Universe the only existing Black Hole?," *ArXiv e-prints*, June 2010.

[14] A. Gregori, "On the Critical Temperatures of Superconductors: a Quantum Gravity Approach," *arXiv e-prints*, July 2010.

[15] A. Gregori, "About combinatorics, and observables," *arXiv:0712.0471*, 2007.

[16] A. Gregori, "Combinatorics, observables, and String Theory," *arXiv:1103.4000*, 2011.

[17] A. Gregori, "Combinatorics, observables, and String Theory, part II," *arXiv:1103.3998*, 2011.

[18] A. Gregori, "Cp violation: beyond field theory?," 2012.

[19] A. Gregori, "A physical universe from the universe of codes," 2012.

[20] A. Gregori, "The superstring representation of the universe of codes," 2012.

[21] A. Gregori, "The spectrum of the universe of codes," 2013.

[22] A. Gregori, "Non-perturbative resonances of the electromagnetic interaction," 2016.

[23] M. Rabinowitz, "n-dimensional gravity: Little black holes, dark matter, and ball lightning," *Int. J. Theor. Phys.*, vol. 40, pp. 875–901, 2001.

[24] J. M. Bardeen, B. Carter, and S. W. Hawking, "The four laws of black hole mechanics," *Commun. Math. Phys.*, vol. 31, pp. 161–170, 1973.

[25] J. D. Bekenstein, "Black holes and entropy," *Phys. Rev.*, vol. D7, pp. 2333–2346, 1973.

[26] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*. Pergamon Press, 1951.

[27] C. Vafa, "Evidence for F-Theory," *Nucl. Phys.*, vol. B469, pp. 403–418, 1996.

[28] L. J. Dixon, V. Kaplunovsky, and J. Louis, “Moduli dependence of string loop corrections to gauge coupling constants,” *Nucl. Phys.*, vol. B355, pp. 649–688, 1991.

[29] E. Kiritsis and C. Kounnas, “Infrared regularization of superstring theory and the one loop calculation of coupling constants,” *Nucl. Phys.*, vol. B442, pp. 472–493, 1995.

[30] E. Kiritsis, C. Kounnas, P. M. Petropoulos, and J. Rizos, “Universality properties of n = 2 and n = 1 heterotic threshold corrections,” *Nucl. Phys.*, vol. B483, pp. 141–171, 1997.

[31] E. Kiritsis, C. Kounnas, P. M. Petropoulos, and J. Rizos, “Solving the decompactification problem in string theory,” *Phys. Lett.*, vol. B385, pp. 87–95, 1996.

[32] E. Kiritsis, C. Kounnas, P. M. Petropoulos, and J. Rizos, “String threshold corrections in models with spontaneously broken supersymmetry,” *Nucl. Phys.*, vol. B540, pp. 87–148, 1999.

[33] A. Gregori and C. Kounnas, “Four-dimensional n = 2 superstring constructions and their (non-)perturbative duality connections,” *Nucl. Phys.*, vol. B560, pp. 135–153, 1999.

[34] E. Kiritsis, N. A. Obers, and B. Pioline, “Heterotic/type ii triality and instantons on k3,” *JHEP*, vol. 01, p. 029, 2000.

[35] I. Antoniadis, E. Gava, and K. S. Narain, “Moduli corrections to gravitational couplings from string loops,” *Phys. Lett.*, vol. B283, pp. 209–212, 1992.

[36] I. Antoniadis, E. Gava, and K. S. Narain, “Moduli corrections to gauge and gravitational couplings in four-dimensional superstrings,” *Nucl. Phys.*, vol. B383, pp. 93–109, 1992.

[37] I. Antoniadis, H. Partouche, and T. R. Taylor, “Duality of n = 2 heterotic-type i compactifications in four dimensions,” *Nucl. Phys.*, vol. B499, pp. 29–44, 1997.

[38] J. A. Harvey and G. Moore, “Algebras, bps states, and strings,” *Nucl. Phys.*, vol. B463, pp. 315–368, 1996.

[39] J. A. Harvey and G. Moore, “Fivebrane instantons and r**2 couplings in n = 4 string theory,” *Phys. Rev.*, vol. D57, pp. 2323–

2328, 1998.

[40] S. Ferrara, J. A. Harvey, A. Strominger, and C. Vafa, "Second quantized mirror symmetry," *Phys. Lett.*, vol. B361, pp. 59–65, 1995.

[41] J. A. Harvey and G. Moore, "Exact gravitational threshold correction in the fhsv model," *Phys. Rev.*, vol. D57, pp. 2329–2336, 1998.

[42] G. Aldazabal, A. Font, L. E. Ibanez, and F. Quevedo, "Heterotic/heterotic duality in d=6,4," *Phys. Lett.*, vol. B380, pp. 33–41, 1996.

[43] A. Sen and C. Vafa, "Dual pairs of type ii string compactification," *Nucl. Phys.*, vol. B455, pp. 165–187, 1995.

[44] E. G. Gimon and C. V. Johnson, "K3 orientifolds," *Nucl. Phys.*, vol. B477, pp. 715–745, 1996.

[45] I. Antoniadis, J. P. Derendinger, and C. Kounnas, "Non-perturbative temperature instabilities in n = 4 strings," *Nucl. Phys.*, vol. B551, pp. 41–77, 1999.

[46] S. Kachru and C. Vafa, "Exact results for n=2 compactifications of heterotic strings," *Nucl. Phys.*, vol. B450, pp. 69–89, 1995.

[47] M. Berkooz and R. G. Leigh, "A d = 4 n = 1 orbifold of type i strings," *Nucl. Phys.*, vol. B483, pp. 187–208, 1997.

[48] I. Antoniadis, E. Gava, and K. S. Narain, "Moduli corrections to gauge and gravitational couplings in four-dimensional superstrings," *Nucl. Phys.*, vol. B383, pp. 93–109, 1992.

[49] A. E. Faraggi, C. Kounnas, S. E. M. Nooij, and J. Rizos, "Towards the classification of z(2) x z(2) fermionic models," 2003.

[50] A. E. Faraggi, C. Kounnas, S. E. M. Nooij, and J. Rizos, "Classification of the chiral z(2) x z(2) fermionic models in the heterotic superstring," *Nucl. Phys.*, vol. B695, pp. 41–72, 2004.

[51] A. E. Faraggi, C. Kounnas, and J. Rizos, "Chiral family classification of fermionic z(2) x z(2) heterotic orbifold models," *Phys. Lett.*, vol. B648, pp. 84–89, 2007.

[52] P. S. Aspinwall and D. R. Morrison, "Point-like instantons on k3 orbifolds," *Nucl. Phys.*, vol. B503, pp. 533–564, 1997.

[53] P. Horava and E. Witten, "Heterotic and type i string dynamics from eleven dimensions," *Nucl. Phys.*, vol. B460, pp. 506–524, 1996.

[54] P. Horava and E. Witten, "Eleven-dimensional supergravity on a manifold with boundary," *Nucl. Phys.*, vol. B475, pp. 94–114, 1996.

[55] R. Dijkgraaf, E. Verlinde, and H. Verlinde, "C=1 conformal fields theories on riemann surfaces," *Comm. Math. Phys.*, vol. 115, p. 2264, 1988.

[56] E. G. Gimon and J. Polchinski, "Consistency conditions for orientifolds and d-manifolds," *Phys. Rev.*, vol. D54, pp. 1667–1676, 1996.

[57] C. Angelantonj, "Comments on open-string orbifolds with a non-vanishing b(ab)," *Nucl. Phys.*, vol. B566, pp. 126–150, 2000.

[58] I. Antoniadis, E. Dudas, and A. Sagnotti, "Supersymmetry breaking, open strings and m-theory," *Nucl. Phys.*, vol. B544, pp. 469–502, 1999.

[59] I. Antoniadis, G. D'Appollonio, E. Dudas, and A. Sagnotti, "Partial breaking of supersymmetry, open strings and m-theory," *Nucl. Phys.*, vol. B553, pp. 133–154, 1999.

[60] CERN press release #25.11, 13 December 2011.

[61] S. Coleman, *Aspects of Symmetry.* Cambridge University Press, 1985.

[62] F. A. Bais, "To be or not to be? Magnetic monopoles in non-Abelian gauge theories," 2004.

[63] M. Tanabashi *et al.*, "The review of particle physics," *Phys. Rev. D*, vol. 98, p. 030001, 2018.

[64] G. Greene *et al.*, "New determination of the deuteron binding energy and the neutron mass"," *Phys. Rev. Lett.*, vol. 56, pp. 819–822, 1986.

[65] Y. Fukuda, T. Hayakawa, E. Ichihara, K. Inoue, K. Ishihara, H. Ishino, Y. Itow, T. Kajita, J. Kameda, S. Kasuga, K. Kobayashi, Y. Kobayashi, Y. Koshio, M. Miura, M. Nakahata, S. Nakayama, A. Okada, K. Okumura, N. Sakurai, M. Shiozawa, Y. Suzuki, Y. Takeuchi, Y. Totsuka, S. Yamada, M. Earl, A. Habig, E. Kearns, M. D. Messier, K. Scholberg, J. L. Stone, L. R. Sulak, C. W. Walter, M. Goldhaber, T. Barszczxak, D. Casper, W. Gajewski, P. G. Halverson, J. Hsu, W. R. Kropp, L. R. Price, F. Reines, M. Smy, H. W. Sobel, M. R. Vagins, K. S. Ganezer, W. E. Keig, R. W. Ellsworth, S. Tasaka, J. W. Flanagan, A. Kibayashi, J. G. Learned, S. Matsuno, V. J. Stenger, D. Takemori, T. Ishii, J. Kanzaki, T. Kobayashi, S. Mine, K. Nakamura, K. Nishikawa, Y. Oyama, A. Sakai, M. Sakuda, O. Sasaki, S. Echigo, M. Kohama, A. T. Suzuki, T. J. Haines, E. Blaufuss, B. K. Kim, R. Sanford, R. Svoboda, M. L. Chen, Z. Conner, J. A. Goodman, G. W. Sullivan, J. Hill, C. K. Jung, K. Martens, C. Mauger, C. McGrew, E. Sharkey, B. Viren, C. Yanagisawa, W. Doki, K. Miyano, H. Okazawa, C. Saji, M. Takahata, Y. Nagashima, M. Takita, T. Yamaguchi, M. Yoshida, S. B. Kim, M. Etoh, K. Fujita, A. Hasegawa, T. Hasegawa, S. Hatakeyama, T. Iwamoto, M. Koga, T. Maruyama, H. Ogawa, J. Shirai, A. Suzuki, F. Tsushima, M. Koshiba, M. Nemoto, K. Nishijima, T. Futagami, Y. Hayato, Y. Kanaya, K. Kaneyuki, Y. Watanabe, D. Kielczewska, R. A. Doyle, J. S. George, A. L. Stachyra, L. L. Wai, R. J. Wilkes, and K. K. Young, "Evidence for oscillation of atmospheric neutrinos," *Physical Review Letters*, vol. 81, pp. 1562–1567, aug 1998.

[66] A. Aguilar-Arevalo, B. Brown, L. Bugel, G. Cheng, J. Conrad, R. Cooper, R. Dharmapalan, A. Diaz, Z. Djurcic, D. Finley, R. Ford, F. Garcia, G. Garvey, J. Grange, E.-C. Huang, W. Huelsnitz, C. Ignarra, R. Johnson, G. Karagiorgi, T. Katori, T. Kobilarcik, W. Louis, C. Mariani, W. Marsh, G. Mills, J. Mirabal, J. Monroe, C. Moore, J. Mousseau, P. Nienaber, J. Nowak, B. Osmanov, Z. Pavlovic, D. Perevalov, H. Ray, B. Roe, A. Russell, M. Shaevitz, J. Spitz, I. Stancu, R. Tayloe, R. Thorn-

ton, M. Tzanov, R. V. de Water, D. White, D. Wickremasinghe, and E. Z. and, "Significant excess of electronlike events in the MiniBooNE short-baseline neutrino experiment," *Physical Review Letters*, vol. 121, nov 2018.

[67] MicroBooNE Collaboration, P. Abratenko, R. An, J. Anthony, L. Arellano, J. Asaadi, A. Ashkenazi, S. Balasubramanian, B. Baller, C. Barnes, G. Barr, V. Basque, L. Bathe-Peters, O. B. Rodrigues, S. Berkman, A. Bhanderi, A. Bhat, M. Bishai, A. Blake, T. Bolton, J. Y. Book, L. Camilleri, D. Caratelli, I. C. Terrazas, F. Cavanna, G. Cerati, Y. Chen, D. Cianci, G. H. Collin, J. M. Conrad, M. Convery, L. Cooper-Troendle, J. I. Crespo-Anadon, M. Del Tutto, S. R. Dennis, P. Detje, A. Devitt, R. Diurba, R. Dorrill, K. Duffy, S. Dytman, B. Eberly, A. Ereditato, L. Escudero-Sanchez, J. J. Evans, R. Fine, G. A. F. Aguirre, R. S. Fitzpatrick, B. T. Fleming, N. Foppiani, D. Franco, A. P. Furmanski, D. Garcia-Gamez, S. Gardiner, G. Ge, V. Genty, S. Gollapinni, O. Goodwin, E. Gramellini, P. Green, H. Greenlee, W. Gu, R. Guenette, P. Guzowski, L. Hagaman, O. Hen, C. Hilgenberg, G. A. Horton-Smith, A. Hourlier, R. Itay, C. James, X. Ji, L. Jiang, J. H. Jo, R. A. Johnson, Y. J. Jwa, D. Kalelo, D. Kalra, N. Kamp, N. Kaneshige, G. Karagiorgi, W. Ketchum, M. Kirby, T. Kobilarcik, I. Kreslo, R. LaZur, I. Lepetic, K. Li, Y. Li, K. Lin, A. Lister, B. R. Littlejohn, W. C. Louis, X. Luo, K. Manivannan, C. Mariani, D. Marsden, J. Marshall, D. A. M. Caicedo, K. Mason, A. Mastbaum, N. McConkey, V. Meddage, T. Mettler, K. Miller, J. Mills, K. Mistry, T. Mohayai, A. Mogan, J. Moon, M. Mooney, A. F. Moor, C. D. Moore, L. M. Lepin, J. Mousseau, M. Murphy, D. Naples, A. Navrer-Agasson, M. Nebot-Guinot, R. K. Neely, D. A. Newmark, J. Nowak, M. Nunes, O. Palamara, V. Paolone, A. Papadopoulou, V. Papavassiliou, S. F. Pate, N. Patel, A. Paudel, Z. Pavlovic, E. Piasetzky, I. Ponce-Pinto, S. Prince, X. Qian, J. L. Raaf, V. Radeka, A. Rafique, M. Reggiani-Guzzo, L. Ren, L. C. J. Rice, L. Rochester, J. R. Rondon, M. Rosenberg, M. Ross-Lonergan, B. Russell, G. Sca-

navini, D. W. Schmitz, A. Schukraft, W. Seligman, M. H. Shaevitz, R. Sharankova, J. Shi, J. Sinclair, A. Smith, E. L. Snider, M. Soderberg, S. Soldner-Rembold, S. R. Soleti, P. Spentzouris, J. Spitz, M. Stancari, J. S. John, T. Strauss, K. Sutton, S. Sword-Fehlberg, A. M. Szelc, W. Tang, K. Terao, M. Thomson, C. Thorpe, D. Totani, M. Toups, Y. T. Tsai, M. A. Uchida, T. Usher, W. Van De Pontseele, B. Viren, M. Weber, H. Wei, Z. Williams, S. Wolbers, T. Wongjirad, M. Wospakrik, K. Wresilo, N. Wright, W. Wu, E. Yandel, T. Yang, G. Yarbrough, L. E. Yates, H. W. Yu, G. P. Zeller, J. Zennamo, and C. Zhang, "Search for an excess of electron neutrino interactions in microboone using multiple final state topologies," 2021.

[68] P. Abratenko, R. An, J. Anthony, L. Arellano, J. Asaadi, A. Ashkenazi, S. Balasubramanian, B. Baller, C. Barnes, G. Barr, V. Basque, L. Bathe-Peters, O. B. Rodrigues, S. Berkman, A. Bhanderi, A. Bhat, M. Bishai, A. Blake, T. Bolton, J. Book, L. Camilleri, D. Caratelli, I. C. Terrazas, F. Cavanna, G. Cerati, Y. Chen, D. Cianci, G. Collin, J. Conrad, M. Convery, L. Cooper-Troendle, J. Crespo-Anadón, M. D. Tutto, S. Dennis, P. Detje, A. Devitt, R. Diurba, R. Dorrill, K. Duffy, S. Dytman, B. Eberly, A. Ereditato, J. Evans, R. Fine, G. F. Aguirre, R. Fitzpatrick, B. Fleming, N. Foppiani, D. Franco, A. Furmanski, D. Garcia-Gamez, S. Gardiner, G. Ge, V. Genty, S. Gollapinni, O. Goodwin, E. Gramellini, P. Green, H. Greenlee, W. Gu, R. Guenette, P. Guzowski, L. Hagaman, O. Hen, C. Hilgenberg, G. Horton-Smith, A. Hourlier, R. Itay, C. James, X. Ji, L. Jiang, J. Jo, R. Johnson, Y.-J. Jwa, D. Kalra, N. Kamp, N. Kaneshige, G. Karagiorgi, W. Ketchum, M. Kirby, T. Kobilarcik, I. Kreslo, I. Lepetic, K. Li, Y. Li, K. Lin, B. Littlejohn, W. Louis, X. Luo, K. Manivannan, C. Mariani, D. Marsden, J. Marshall, D. M. Caicedo, K. Mason, A. Mastbaum, N. McConkey, V. Meddage, T. Mettler, K. Miller, J. Mills, K. Mistry, A. Mogan, T. Mohayai, J. Moon, M. Mooney, A. Moor, C. Moore, L. M. Lepin, J. Mousseau, M. Murphy, D. Naples, A. Navrer-Agasson, M. Nebot-Guinot, R. Neely, D. Newmark,

J. Nowak, M. Nunes, O. Palamara, V. Paolone, A. Papadopoulou, V. Papavassiliou, S. Pate, N. Patel, A. Paudel, Z. Pavlovic, E. Piasetzky, I. Ponce-Pinto, S. Prince, X. Qian, J. Raaf, V. Radeka, A. Rafique, M. Reggiani-Guzzo, L. Ren, L. Rice, L. Rochester, J. R. Rondon, M. Rosenberg, M. Ross-Lonergan, G. Scanavini, D. Schmitz, A. Schukraft, W. Seligman, M. Shaevitz, R. Sharankova, J. Shi, J. Sinclair, A. Smith, E. Snider, M. Soderberg, S. Söldner-Rembold, P. Spentzouris, J. Spitz, M. Stancari, J. John, T. Strauss, K. Sutton, S. Sword-Fehlberg, A. Szelc, W. Tang, K. Terao, C. Thorpe, D. Totani, M. Toups, Y.-T. Tsai, M. Uchida, T. Usher, W. V. D. Pontseele, B. Viren, M. Weber, H. Wei, Z. Williams, S. Wolbers, T. Wongjirad, M. Wospakrik, K. Wresilo, N. Wright, W. Wu, E. Yandel, T. Yang, G. Yarbrough, L. Yates, H. Yu, G. Zeller, J. Zennamo, and C. Z. and, "Search for an anomalous excess of charged-current quasielastic ν_e interactions with the MicroBooNE experiment using deep-learning-based reconstruction," *Physical Review D*, vol. 105, jun 2022.

[69] MicroBooNE Collaboration, P. Abratenko, R. An, J. Anthony, L. Arellano, J. Asaadi, A. Ashkenazi, S. Balasubramanian, B. Baller, C. Barnes, G. Barr, V. Basque, L. Bathe-Peters, O. B. Rodrigues, S. Berkman, A. Bhanderi, A. Bhat, M. Bishai, A. Blake, T. Bolton, J. Y. Book, L. Camilleri, D. Caratelli, I. C. Terrazas, F. Cavanna, G. Cerati, Y. Chen, D. Cianci, J. M. Conrad, M. Convery, L. Cooper-Troendle, J. I. Crespo-Anadon, M. Del Tutto, S. R. Dennis, P. Detje, A. Devitt, R. Diurba, R. Dorrill, K. Duffy, S. Dytman, B. Eberly, A. Ereditato, L. Escudero-Sanchez, J. J. Evans, R. Fine, G. A. F. Aguirre, R. S. Fitzpatrick, B. T. Fleming, N. Foppiani, D. Franco, A. P. Furmanski, D. Garcia-Gamez, S. Gardiner, G. Ge, S. Gollapinni, O. Goodwin, E. Gramellini, P. Green, H. Greenlee, W. Gu, R. Guenette, P. Guzowski, L. Hagaman, O. Hen, C. Hilgenberg, G. A. Horton-Smith, A. Hourlier, R. Itay, C. James, X. Ji, L. Jiang, J. H. Jo, R. A. Johnson, Y. J. Jwa, D. Kalra, N. Kamp, N. Kaneshige, G. Karagiorgi, W. Ketchum, M. Kirby, T. Kobi-

larcik, I. Kreslo, R. LaZur, I. Lepetic, K. Li, Y. Li, K. Lin, A. Lister, B. R. Littlejohn, W. C. Louis, X. Luo, K. Manivannan, C. Mariani, D. Marsden, J. Marshall, D. A. M. Caicedo, K. Mason, A. Mastbaum, N. McConkey, V. Meddage, T. Mettler, K. Miller, J. Mills, K. Mistry, T. Mohayai, A. Mogan, J. Moon, M. Mooney, A. F. Moor, C. D. Moore, L. M. Lepin, J. Mousseau, M. Murphy, D. Naples, A. Navrer-Agasson, M. Nebot-Guinot, R. K. Neely, D. A. Newmark, J. Nowak, M. Nunes, O. Palamara, V. Paolone, A. Papadopoulou, V. Papavassiliou, S. F. Pate, N. Patel, A. Paudel, Z. Pavlovic, E. Piasetzky, I. Ponce-Pinto, S. Prince, X. Qian, J. L. Raaf, V. Radeka, A. Rafique, M. Reggiani-Guzzo, L. Ren, L. C. J. Rice, L. Rochester, J. R. Rondon, M. Rosenberg, M. Ross-Lonergan, G. Scanavini, D. W. Schmitz, A. Schukraft, W. Seligman, M. H. Shaevitz, R. Sharankova, J. Shi, J. Sinclair, A. Smith, E. L. Snider, M. Soderberg, S. Soldner-Rembold, S. R. Soleti, P. Spentzouris, J. Spitz, M. Stancari, J. S. John, T. Strauss, K. Sutton, S. Sword-Fehlberg, A. M. Szelc, W. Tang, K. Terao, M. Thomson, C. Thorpe, D. Totani, M. Toups, Y. T. Tsai, M. A. Uchida, T. Usher, W. Van De Pontseele, B. Viren, M. Weber, H. Wei, Z. Williams, S. Wolbers, T. Wongjirad, M. Wospakrik, K. Wresilo, N. Wright, W. Wu, E. Yandel, T. Yang, G. Yarbrough, L. E. Yates, H. W. Yu, G. P. Zeller, J. Zennamo, and C. Zhang, "Search for an anomalous excess of charged-current ν_e interactions without pions in the final state with the microboone experiment," 2021.

[70] MicroBooNE Collaboration, P. Abratenko, R. An, J. Anthony, L. Arellano, J. Asaadi, A. Ashkenazi, S. Balasubramanian, B. Baller, C. Barnes, G. Barr, V. Basque, L. Bathe-Peters, O. B. Rodrigues, S. Berkman, A. Bhanderi, A. Bhat, M. Bishai, A. Blake, T. Bolton, J. Y. Book, L. Camilleri, D. Caratelli, I. C. Terrazas, F. Cavanna, G. Cerati, Y. Chen, D. Cianci, J. M. Conrad, M. Convery, L. Cooper-Troendle, J. I. Crespo-Anadon, M. Del Tutto, S. R. Dennis, P. Detje, A. Devitt, R. Diurba, R. Dorrill, K. Duffy, S. Dytman, B. Eberly, A. Ereditato, J. J.

Evans, R. Fine, G. A. F. Aguirre, R. S. Fitzpatrick, B. T. Fleming, N. Foppiani, D. Franco, A. P. Furmanski, D. Garcia-Gamez, S. Gardiner, G. Ge, S. Gollapinni, O. Goodwin, E. Gramellini, P. Green, H. Greenlee, W. Gu, R. Guenette, P. Guzowski, L. Hagaman, O. Hen, C. Hilgenberg, G. A. Horton-Smith, A. Hourlier, R. Itay, C. James, X. Ji, L. Jiang, J. H. Jo, R. A. Johnson, Y. J. Jwa, D. Kalra, N. Kamp, N. Kaneshige, G. Karagiorgi, W. Ketchum, M. Kirby, T. Kobilarcik, I. Kreslo, I. Lepetic, K. Li, Y. Li, K. Lin, B. R. Littlejohn, W. C. Louis, X. Luo, K. Manivannan, C. Mariani, D. Marsden, J. Marshall, D. A. M. Caicedo, K. Mason, A. Mastbaum, N. McConkey, V. Meddage, T. Mettler, K. Miller, J. Mills, K. Mistry, T. Mohayai, A. Mogan, J. Moon, M. Mooney, A. F. Moor, C. D. Moore, L. M. Lepin, J. Mousseau, M. Murphy, D. Naples, A. Navrer-Agasson, M. Nebot-Guinot, R. K. Neely, D. A. Newmark, J. Nowak, M. Nunes, O. Palamara, V. Paolone, A. Papadopoulou, V. Papavassiliou, S. F. Pate, N. Patel, A. Paudel, Z. Pavlovic, E. Piasetzky, I. Ponce-Pinto, S. Prince, X. Qian, J. L. Raaf, V. Radeka, A. Rafique, M. Reggiani-Guzzo, L. Ren, L. C. J. Rice, L. Rochester, J. R. Rondon, M. Rosenberg, M. Ross-Lonergan, B. Russell, G. Scanavini, D. W. Schmitz, A. Schukraft, W. Seligman, M. H. Shaevitz, R. Sharankova, J. Shi, J. Sinclair, A. Smith, E. L. Snider, M. Soderberg, S. Soldner-Rembold, P. Spentzouris, J. Spitz, M. Stancari, J. S. John, T. Strauss, K. Sutton, S. Sword-Fehlberg, A. M. Szelc, W. Tang, K. Terao, C. Thorpe, D. Totani, M. Toups, Y. T. Tsai, M. A. Uchida, T. Usher, W. Van De Pontseele, B. Viren, M. Weber, H. Wei, Z. Williams, S. Wolbers, T. Wongjirad, M. Wospakrik, K. Wresilo, N. Wright, W. Wu, E. Yandel, T. Yang, G. Yarbrough, L. E. Yates, H. W. Yu, G. P. Zeller, J. Zennamo, and C. Zhang, "Search for an anomalous excess of inclusive charged-current ν_e interactions in the microboone experiment using wire-cell reconstruction," 2021.

[71] A. A. Aguilar-Arevalo, A. O. Bazarko, S. J. Brice, B. C. Brown, L. Bugel, J. Cao, L. Coney, J. M. Conrad, D. C. Cox, A. Curioni,

Z. Djurcic, D. A. Finley, B. T. Fleming, R. Ford, F. G. Garcia, G. T. Garvey, C. Green, J. A. Green, T. L. Hart, E. Hawker, R. Imlay, R. A. Johnson, P. Kasper, T. Katori, T. Kobilarcik, I. Kourbanis, S. Koutsoliotas, E. M. Laird, J. M. Link, Y. Liu, Y. Liu, W. C. Louis, K. B. M. Mahn, W. Marsh, P. S. Martin, G. McGregor, W. Metcalf, P. D. Meyers, F. Mills, G. B. Mills, J. Monroe, C. D. Moore, R. H. Nelson, P. Nienaber, S. Ouedraogo, R. B. Patterson, D. Perevalov, C. C. Polly, E. Prebys, J. L. Raaf, H. Ray, B. P. Roe, A. D. Russell, V. Sandberg, R. Schirato, D. Schmitz, M. H. Shaevitz, F. C. Shoemaker, D. Smith, M. Sorel, P. Spentzouris, I. Stancu, R. J. Stefanski, M. Sung, H. A. Tanaka, R. Tayloe, M. Tzanov, R. V. de Water, M. O. Wascko, D. H. White, M. J. Wilking, H. J. Yang, G. P. Zeller, and E. D. Zimmerman, "Search for Electron Neutrino Appearance at the $\Delta m^2 \sim 1eV^2$ scale," *Physical Review Letters*, vol. 98, jun 2007.

[72] T. Aaltonen *et al.*, "Measurement of cp-violating asymmetries in $d^0 \to \pi^+\pi^-$ and $d^0 \to k^+k^-$ decays at cdf," *arXiv:1111.5023v1 [hep-ex]*, 2011.

[73] R. Aaij *et al.*, "Evidence for cp violation in time-integrated $d^0 \to h^-h^+$ decay rates," *Phys. Rev. Lett.*, vol. 108, p. 111602, 2012.

[74] R. Aaij *et al.*, "Measurement of cp asymmetry in $d^0 \to k^+k^-$ and $d^0 \to \pi^-\pi^+$ decays," *JHEP*, vol. 07, p. 041, 2014.

[75] N. Dash, "Cp violation in $d^0 \to k_s^0 k_s^0$ decay at belle," *arXiv:1711.10419*, 2017. XXII DAE-BRNS HEP symposium proceeding, submitted by springer to European Physics Journal C.

[76] S. Burles, K. M. Nollet, and M. S. Turner, "Big-bang nucleosynthesis predictions for precision cosmology," *Astrophys. J. 552*, pp. L1–L6, 2001.

[77] Y. Nir, "Cp violation - a new era," 2001.

[78] L. Taylor, "Observation of a new particle with a mass of 125 gev," *CMS Public Website, CERN*, 4 July 2012.

[79] C. Collaboration, “Observation of a new boson with a mass near 125 gev,” *Cms-Pas-Hig-12-020*, 9 July 2012.

[80] A. Rajaraman, T. M. P. Tait, and D. Whiteson, “Two lines or not two lines? that is the question of gamma ray spectra,” *JCAP*, vol. 1209, p. 003, 2012.

[81] T. Bringmann, X. Huang, A. Ibarra, S. Vogl, and C. Weniger, “Fermi lat search for internal bremsstrahlung signatures from dark matter annihilation,” *JCAP*, vol. 1207, p. 054, 2012.

[82] C. Weniger, “A tentative gamma-ray line from dark matter annihilation at the fermi large area telescope,” *JCAP*, vol. 1208, p. 007, 2012.

[83] Y. Li and Q. Yuan, “Testing the 130 gev gamma-ray line with high energy resolution detectors,” *Phys. Lett. B*, vol. 715, pp. 35–37, 2012.

[84] M. Su and D. P. Finkbeiner, “Strong evidence for gamma-ray line emission from the inner galaxy,”

[85] A. A. Penzias and R. W. Wilson, “A measurement of excess antenna temperature at 4080 mc/s,” *Astrophysical Journal*, vol. 142, p. 414, 1965.

[86] J. C. Mather, D. J. Fixsen, R. A. Shafer, C. Mosier, and D. T. Wilkinson, “Calibrator design for the cobe far infrared absolute spectrophotometer (firas),” *Astrophys. J.*, vol. 512, pp. 511–520, 1999.

[87] D. Clowe *et al.*, “A direct empirical proof of the existence of dark matter,” *Astrophys. J.*, vol. 648, pp. L109–L113, 2006.

[88] W. K. Ford-Jr and V. Rubin, “Rotation of the andromeda nebula from a spectroscopic survey of emission regions,” *Astrophysical Journal*, vol. 159, p. 379, 1970.

[89] W. K. Ford-Jr, V. Rubin, and N. Thonnard, “Rotational properties of 21 sc galaxies with a large range of luminosities and radii from ngc 4605 (r=4kpc) to ugc 2885 (r=122kpc),” *Astrophysical Journal*, vol. 238, p. 471, 1980.

[90] J. K. Webb, V. V. Flambaum, C. W. Churchill, M. J. Drinkwa-

ter, and J. D. Barrow, "Evidence for time variation of the fine structure constant," *Phys. Rev. Lett.*, vol. 82, pp. 884–887, 1999.

[91] V. A. Dzuba, V. V. Flambaum, and J. K. Webb, "Calculations of the relativistic effects in many-electron atoms and space-time variation of fundamental constants," *Phys. Rev.*, vol. A59, pp. 230–237, 1999.

[92] J. K. Webb *et al.*, "Further evidence for cosmological evolution of the fine structure constant," *Phys. Rev. Lett.*, vol. 87, p. 091301, 2001.

[93] M. T. Murphy *et al.*, "Possible evidence for a variable fine structure constant from qso absorption lines: motivations, analysis and results," *Mon. Not. Roy. Astron. Soc.*, vol. 327, p. 1208, 2001.

[94] X. Calmet and H. Fritzsch, "Symmetry breaking and time variation of gauge couplings," *Phys. Lett.*, vol. B540, pp. 173–178, 2002.

[95] X. Calmet and H. Fritzsch, "The cosmological evolution of the nucleon mass and the electroweak coupling constants," *Eur. Phys. J.*, vol. C24, pp. 639–642, 2002.

[96] V. V. Flambaum and E. V. Shuryak, "Limits on cosmological variation of strong interaction and quark masses from big bang nucleosynthesis, cosmic, laboratory and oklo data," *Phys. Rev.*, vol. D65, p. 103503, 2002.

[97] T. Damour and F. Dyson, "The oklo bound on the time variation of the fine-structure constant revisited," *Nucl. Phys.*, vol. B480, pp. 37–54, 1996.

[98] A. Dabrincourt-Malassé, "Nouveau regard sur l'origine de l'homme," *La Recherche*, vol. 286, pp. 45–51, 1996.

[99] A. Dabrincourt-Malassé, "L'hominisation et la théorie des systèmes dynamiques non linéaires," *Revue de biologie mathématique*, vol. 286, pp. 117–119, 1992.

[100] A. Dabrincourt-Malassé, "Continuity and discontinuity during modalities of hominization," *Quaternary International*, vol. 19,

1993.

[101] A. Dabrincourt-Malassé, “Modeling of cranio-facial architecture during ontogenesis and phylogenesis,” in *The Head-Neck sensory motor system*, Oxford University Press, New-York-Oxford, 1992.

[102] C.-M. Chang, A. H. C. Neto, and A. R. Bishop, “Mutagenesis and metallic dna,” 2000.

[103] L. Frappat, A. Sciarrino, and P. Sorba, “A crystal base for the genetic code,” *Physics Letters A*, vol. 250, p. 214, 1998.

[104] L. Frappat, A. Sciarrino, and P. Sorba, “Crystalizing the genetic code,” *Journal of biological physics*, vol. 27 1, pp. 1–34, 2001.

[105] H. Diamant and D. Andelman, “Binding of molecules to dna and other semiflexible polymers,” *Physical Review E*, vol. 61, p. 6740, 2000.

[106] V. L. Golo and Y. S. Volkov, “Tautomeric transitions in dna,” 2001.

[107] P. R. Renne, A. L. Deino, F. J. Hilgen, K. F. Kuiper, D. F. Mark, W. S. Mitchell, L. E. Morgan, R. Mundil, and J. Smit, “Time Scales of Critical Events Around the Cretaceous-Paleogene Boundary,” *Science*, vol. 339, pp. 684–687, 2013.

[108] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, “Theory of superconductivity,” *Phys. Rev.*, vol. 108, p. 1175, 1957.

[109] M. Tinkham, *Introduction to Superconductivity.* Dover Publications, 2004.

[110] H. Takahashi, K. Igawa, K. Arii, Y. Kamihara, M. Hirano, and H. Hosono, “Superconductivity at 43 K in an iron-based layered compound LaO1-xFxFeAs,” *Nature*, vol. 453, pp. 376–378, 2008.

[111] R. Zhi-An *et al.*, “Superconductivity at 43 K in an iron-based layered compound LaO1-xFxFeAs,” *Chinese Phys. Lett.*, vol. 25, pp. 2215–2216, 2008.

[112] R. Zhi-An *et al.*, “Fluorinated superconductors ,” *NPG Asia Materials*, Published online 14 July 2008.

[113] M. K. Wu, J. R. Ashburn, C. J. Torng, P. H. Hor, R. L. Meng,

L. Gao, Z. J. Huang, Y. Q. Wang, and C. W. Chu, "Superconductivity at 93 k in a new mixed-phase y-ba-cu-o compound system at ambient pressure," *Phys. Rev. Lett.*, vol. 58, pp. 908–910, Mar 1987.

[114] H. Maeda, Y. Tanaka, M. Fukutomi, and T. Asano, "A new high-t_c oxide superconductor without a rare earth element," *Jpn. J. Appl. Phys.*, vol. 27, pp. L209–L210, 1988.

[115] Z. Sheng, A. Hermann, A. El Ali, C. Almasan, J. Estrada, T. Datta, and R. Matson, "Superconductivity at 90 K in the Tl-Ba-Cu-O system," *Physical Review Letters*, vol. 60, p. 937, 1988.

[116] S. S. P. Parkin, V. Y. Lee, A. I. Nazzal, R. Savoy, R. Beyers, and S. J. La Placa, "A new class of crystal structures exhibiting volume superconductivity at up to 110 k," *Phys. Rev. Lett.*, vol. 61, pp. 750–753, 1988.

[117] N. Khare, ed., *Handbook of High-Temperature Superconductor Electronics*. Marcel Dekker, 2003.

[118] Z. Sheng and A. M. Hermann, "Superconductivity in the rare-earth-free Tl-Ba-Cu-O system above liquid-nitrogen temperature," *Nature*, vol. 332, p. 55, 1988.

[119] S. N. Putilin, E. V. Antipov, O. Chmaissem, and M. Marezio, "Superconductivity at 94 K in HgBa2Cu04+δ," *Nature*, vol. 362, p. 226, 1993.

[120] J. L. Wagner, T. M. Clemens, D. C. Mathew, O. Chmaissem, B. Dabrowski, J. Jorgensen, and D. Hinks, "Universal phase diagrams and ideal high temperature superconductors: Hgba2cuo4+δ," in *Phase Transitions and Self-Organization in Electronic and Molecular Networks* (M. F. Thorpe and J. C. Phillips, eds.), pp. 331–339, Springer US, 2002.

[121] C. W. Chu, L. Gao, F. Chen, Z. J. Huang, R. L. Meng, and Y. Y. Xue, "Superconductivity above 150 k in hgba2ca2cu3o8+δ at high pressures," *Nature*, vol. 365, p. 323, 1993.

[122] J. Eck, "Superconductors," *http://www.superconductors.org/*.