1. Introduction

Our knowledge of spacetime is described by two existing theories:

- General Relativity
- The Standard Model of particle physics

General relativity describes spacetime as far as large scales are concerned and is based on the geometric paradigm discovered by Riemann. It replaces the flat (pseudo) metric of Poincaré, Einstein, and Minkowski, by a curved spacetime metric whose components form the gravitational potential. The basic equations are Einstein equations (Figure 1) which have a clear geometric meaning and are derived from a simple action principle. Many processes in physics can be understood in terms of an action principle, which says, roughly speaking, that the actual observed process minimizes some functional, the action, over the space of possible processes. A simple example is Fermat’s principle in optics which asserts that light of a given frequency traverses the path between two points which takes the least time. In the case of Einstein’s equation the action is a functional on the set of all possible spacetime configurations and is computed by integrating the spacetime curvature over all of spacetime.
This Einstein-Hilbert action $S_{EH}$, from which the Einstein equations are derived in empty space, is replaced in the presence of matter by the combination

$$S = S_{EH} + S_{SM}$$

where the second term $S_{SM}$ is the standard model action which encapsulates our knowledge of all the different kinds of elementary particles to be found in nature. These are typically discovered by smashing together highly accelerated beams of other particles in huge particle accelerators such as at CERN (Figure 2) and what one finds, even at the kinds of energies observable today (about 100 GeV) is a veritable zoo of particles – electrons, quarks, neutrinos of various flavors to name a few – and their antiparticles as well as the force carriers (cf. Figure 4). The action $S_{SM}$ encodes what is known so far about this array of particles and in principle describes all the matter and particles of force in the universe, with the exception of gravity, to our best current understanding. It also has some twenty or more parameters put in by hand according to the experimental data, such as the masses of the various particles. The values of these parameters is a great mystery.

While the Einstein-Hilbert action $S_{EH}$ has a clear geometric meaning, the additional term $S_{SM}$ (cf. Figure 3) is quite complicated (it takes about four hours to typeset the formula) and
Figure 3. Standard Model Lagrangian

is begging for a better understanding— in the words of an old friend\textsuperscript{1}, it is a Shakespearean king disguised as a beggar\textsuperscript{2}.

Our goal in this expository text is to explain that a conceptual understanding of the full action functional is now available (joint work with A. Chamseddine and M. Marcolli \cite{Chamseddine:2001gq, Chamseddine:2003vz}) and shows that the additional term $S_M$ exhibits the fine texture of the geometry of spacetime. This fine texture appears as the product of the ordinary 4-dimensional continuum by a very specific finite discrete space $F$. Just to get a mental picture one may, in first approximation,

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{2}I told him “I hope the beggar has diamonds in its pockets”
\end{itemize}
\end{footnotesize}
think of $F$ as a space consisting of two points. The product space then appears as a 4-dimensional continuum with “two sides”. As we shall see, after a judicious choice of $F$, one obtains the full action functional (I) as describing pure gravity on the product space $M \times F$. It is crucial, of course, to understand the “raison d’être” of the space $F$ and to explain why crossing the ordinary continuum with such a space is necessary from first principles. As we shall see below such an explanation is now available (cf. [8]).

2. THE QUANTUM FORMALISM AND ITS INFINITESIMAL VARIABLES

Let us start by explaining the formalism of quantum mechanics and how it provides a natural home for the notion of infinitesimal variable which is at the heart of the beginning of the “calculus”.

One essential difference between the way Newton and Leibniz treated infinitesimals is that for Newton an infinitesimal is a variable. More precisely, according to Newton:\footnote{See A. N. Krylov Leonhard Euler. Talk given on October 5 1933, translated by N. G. Kuznetsov.}

\begin{quote}
In a certain problem, a variable is the quantity that takes an infinite number of values which are quite determined by this problem and are arranged in a definite order.
\end{quote}

Each of these values is called a particular value of the variable.
A variable is called “infinitesimal” if among its particular values one can be found such that this value itself and all following it are smaller in absolute value than an arbitrary given number.

In the usual mathematical formulation of variables as maps from a set X to the real numbers R, the set X has to be uncountable if some variable has continuous range. But then for any other variable with countable range some of the multiplicities are infinite. This means that discrete and continuous variables cannot coexist in this modern formalism. Fortunately everything is fine and this problem of treating continuous and discrete variables on the same footing is completely solved using the formalism of quantum mechanics which we now briefly recall.

At the beginning of the twentieth century a wealth of experimental data was being collected on the spectra of various chemical elements. These spectra obey experimentally discovered laws, the most notable being the Ritz-Rydberg combination principle (Figure 5). The principle can be stated as follows; spectral lines are indexed by pairs of labels. The statement of the principle then is that certain pairs of spectral lines, when expressed in terms of frequencies, do add up to give another line in the spectrum. Moreover, this happens precisely when the labels are of the form \(i, j\) and \(j, k\), i.e. when the second label \(j\) of the first spectral line coincides with the first label of the second spectral line.

**Figure 5.** Spectral lines and the Ritz-Rydberg law.
In his seminal paper\(^4\) of 1925, Heisenberg showed that this experimental law in fact dictates the algebraic rules for the observable quantities of the microscopic system given by an atom. It was Born who realized that what Heisenberg described in his paper corresponded to replacing classical coordinates with coordinates which no longer commute, but which obey the laws of matrix multiplication. In his own words reported in B.L. van der Waerden’s book\(^5\):

After having sent Heisenberg’s paper to the Zeitschrift für Physik for publication, I began to ponder about his symbolic multiplication, and was soon so involved in it that I thought the whole day and could hardly sleep at night. For I felt there was something fundamental behind it ... And one morning ... I suddenly saw light: Heisenberg’s symbolic multiplication was nothing but the matrix calculus.

In other words, the Ritz–Rydberg law gives the groupoid law:

\[(i, j) \circ (j, k) = (i, k)\]

and the algebra of observables is given by the matrix product

\[(AB)_{jk} = \sum_j A_{ij} B_{jk},\]

for which in general commutativity is lost:

\[(2)\]

\[AB \neq BA.\]

This viewpoint on quantum mechanics was later somewhat obscured by the advent of the Schrödinger equation. The Schrödinger approach shifted the emphasis back to the more traditional technique of solving partial differential equations, while the more modern viewpoint of Heisenberg implied a much more serious change of paradigm, affecting our most basic understanding of the notion of space. We shall assume that the reader has some familiarity with the algebra of matrices of finite dimension and also has some idea of what Hilbert space is: an infinite dimensional complex linear space with a positive inner product, while matrices are replaced in this infinite dimensional set-up by linear operators.

Thus in quantum mechanics the basic change of paradigm has to do with the observable quantities which one would classically describe as real valued functions on a set \(i.e.\) as a map from this set to real numbers. It replaces this classical notion of a “real variable” by the following substitute: a self-adjoint operator in Hilbert space. Note that, if we were dealing with finite dimensional Hilbert spaces a real variable would simply be a Hermitian matrix\(^6\), but to have more freedom, it is more convenient to take the Hilbert space to be the unique separable infinite dimensional Hilbert space. All the usual attributes of real variables such as their range, the number of times a real number is reached as a value of the variable etc... have a perfect analogue in this quantum mechanical setting. The range is the spectrum of the operator, and the spectral multiplicity gives the number of times a real number is reached.

In the early times of quantum mechanics, physicists had a clear intuition of this analogy between operators in Hilbert space (which they called \(q\)-numbers) and variables.

---


\(^6\) A matrix \((A_{ij})\) is Hermitian when \(A_{ji} = \overline{A}_{ij}\) for all \(i, j\).
What is surprising is that the new set-up immediately provides a natural home for the “infinitesimal variables”.

Indeed it is perfectly possible for an operator to be “smaller than \( \epsilon \) for any \( \epsilon \)” without being zero. This happens when the size of the restriction of the operator to subspaces of finite codimension tends to zero when these subspaces decrease (under the natural filtration by inclusion). The corresponding operators are called “compact” and they share with naive infinitesimals all the expected algebraic properties. In fact they comply exactly with Newton’s definition of infinitesimal variables since the list of their characteristic values is a sequence decreasing to 0. All the expected algebraic rules such as infinitesimal \( \times \) bounded = infinitesimal etc. are fulfilled. The only property of the naive infinitesimal calculus that needs to be dropped is the commutativity.

It is only because one drops commutativity that variables with continuous range can coexist with variables with countable range. In the classical formulation of variables, as maps from a set \( X \) to the real numbers, we saw above that discrete variables cannot coexist with continuous variables. The uniqueness of the separable infinite dimensional Hilbert space cures that problem, and variables with continuous range coexist happily with variables with countable range, such as the infinitesimal ones. The only new fact is that they do not commute.

One way to understand the transition from the commutative to the noncommutative is that in the latter case one needs to care about the ordering of the letters when one is writing, whereas the commutative rule oversimplifies the computations. As explained above, it is Heisenberg who discovered that such care was needed when dealing with the coordinates on the phase space of microscopic systems.

At the philosophical level there is something quite satisfactory in the variability of the quantum mechanical observables. Usually when pressed to explain what is the cause of the variability in the external world, the answer that comes naturally to the mind is just: the passing of time. But precisely the quantum world provides a more subtle answer since the reduction of the wave packet which happens in any quantum measurement is nothing else but the replacement of a “q-number” by an actual number which is chosen among the elements in its spectrum. Thus there is an intrinsic “variability” in the quantum world which is so far not reducible to anything classical. The results of observations are intrinsically variable quantities, and this to the point that their values cannot be reproduced from one experiment to the next, but which, when taken altogether, form a q-number.

Heisenberg’s discovery shows that the phase-space of microscopic systems is noncommutative inasmuch as the coordinates on that space no longer satisfy the commutative rule of ordinary algebra. This example of the phase space can be regarded as the historic origin of noncommutative geometry. But what about spacetime itself? We now show why it is a natural step to pass from a commutative spacetime to a noncommutative one.

---

7 The size \( ||T|| \) of an operator is the square root of the size of the spectrum of \( T^*T \).

8 Since for instance the Hilbert space \( L^2([0,1]) \) is the same as \( \ell^2(\mathbb{N}) \).

9 As an example use the “commutative rule” to simplify the following cryptic message I received from a friend : “Je suis alenconnais, et non alsacien. Si t’as besoin d’un conseil nana, je t’attends au coin annales. Qui suis-je? ”
3. Why should spacetime be noncommutative

The full action \( S = S_{EH} + S_{SM} \) of gravity coupled with matter admits a huge natural group of symmetries. The group of invariance for the Einstein-Hilbert action \( S_{EH} \) is the group of diffeomorphisms of the manifold \( M \) and the invariance of the action \( S_{EH} \) is simply the manifestation of its geometric nature. A diffeomorphism acts by permutations of the points of \( M \) so that points have no absolute meaning. The full group \( U \) of invariance of the action (1) is however richer than the group of diffeomorphisms of the manifold \( M \) since one needs to include something called “the group of gauge transformations” which physicists have identified as the symmetry of the matter part \( S_{SM} \). This is defined as the group of maps from the manifold \( M \) to some fixed other group, \( G \), called the ‘gauge group’, which as far as we known is:

\[
G = U(1) \times SU(2) \times SU(3).
\]

The group of diffeomorphisms acts on the group of gauge transformations by permutations of the points of \( M \) and the full group of symmetries of the action \( S \) is the semi-direct product of the two groups (in the same way, the Poincaré group which is the invariance group of special relativity, is the semi-direct product of the group of translations by the group of Lorentz transformations). In particular it is not a simple group\(^{10}\) but is a “composite” and contains a huge normal subgroup.

Now that we know the invariance group \( U \) of the action (1), it is natural to try and find a space \( X \) whose group of diffeomorphisms is simply that group, so that we could hope to interpret the full action as pure gravity on \( X \). This is the old Kaluza-Klein idea. Unfortunately this search is bound to fail if one looks for an ordinary manifold since by a mathematical result, the connected component of the identity in the group of diffeomorphisms is always a simple group, excluding a semi-direct product structure as that of \( U \).

But noncommutative spaces of the simplest kind readily give the answer, modulo a few subtle points. To understand what happens note that for ordinary manifolds the algebraic object corresponding to a diffeomorphism is just an automorphism of the algebra of coordinates i.e. a transformation of the coordinates that does not destroy their algebraic relations. When an algebra is not commutative there is an easy way to construct automorphisms. One takes an element \( u \) of the algebra and one assumes that \( u \) is invertible i.e. its inverse \( u^{-1} \) exists, solution of the equation \( uu^{-1} = u^{-1}u = 1 \). Using \( u \) one obtains an automorphism called inner, by the formula

\[
\alpha(x) = u x u^{-1}, \quad \forall x \in A.
\]

Note that in the commutative case this formula just gives the identity automorphism (since one could then permute \( x \) and \( u^{-1} \)). Thus this construction is interesting only in the noncommutative case. Moreover the inner automorphisms form a subgroup \( \text{Int}(A) \) which is always a normal subgroup of the group of automorphisms\(^{11}\).

\(^{10}\)A simple group is one which cannot be decomposed into smaller pieces, a bit like a prime number cannot be factorized into a product of smaller numbers.

\(^{11}\)Since we deal with involutive algebras we use unitaries: \( u^{-1} = u^* \) to define inner automorphisms.
In the simplest example, where we take for $\mathcal{A}$ the algebra of smooth maps from a manifold $M$ to the algebra of $n \times n$ matrices of complex numbers, one shows that the group Int($\mathcal{A}$) in that case is (locally) isomorphic to the group of gauge transformations i.e. of smooth maps from $M$ to the gauge group $G = \text{PSU}(n)$ (quotient of SU($n$) by its center). Moreover the relation between inner automorphisms and all automorphisms becomes identical to the exact sequence governing the structure of the group $\mathcal{U}$.

It is quite striking that the terminology coming from physics: internal symmetries agrees so well with the mathematical one of inner automorphisms. In the general case only automorphisms that are unitarily implemented in Hilbert space will be relevant but modulo this subtlety one can see at once from the above example the advantage of treating noncommutative spaces on the same footing as the ordinary ones. The next step is to properly define the notion of metric for such spaces and we shall first indulge in a short historical description of the evolution of the definition of the “unit of length” in physics. This will prepare the ground for the introduction to the spectral paradigm of noncommutative geometry in the following section.

4. A BRIEF HISTORY OF THE METRIC SYSTEM

The next step is to understand what is the replacement of the Riemannian paradigm for noncommutative spaces. To prepare for that we now tell the story of the change of paradigm that already took place in the metric system with the replacement of the concrete “mètre étalon” by a spectral unit of measurement.
We describe the corresponding mathematical paradigm of noncommutative geometry in §5. The notion of geometry is intimately tied up with the measurement of length. In the real world such measurement depends on the chosen system of units and the story of the most commonly used system—the metric system—illustrates the difficulties attached to reaching some agreement on a physical unit of length which would unify the previous numerous existing choices.

In 1791 the French Academy of Sciences agreed on the definition of the unit of length in the metric system, the “mètre”, as being \(10^{-7}\) times the quarter of the meridian of the earth (Figure 6). The idea was to measure the length of the arc of the meridian from Barcelone to Dunkerque while the corresponding angle (approximately 9.5°) was determined using the measurement of latitude from reference stars. In a way this was just a refinement of what Eratosthenes had done in Egypt, 250 years BC, to measure the size of the earth (with a precision of 0.4 percent). Thus in 1792 two expeditions were sent to measure this arc of the meridian, one for the Northern portion was led by Delambre and the other for the southern portion was led by Méchain. Both of them were astronomers who were using a new instrument for measuring angles, invented by Borda, a French physicist. The method they used is the method of triangulation (Figure 7) and of concrete measurement of the “base” of one triangle. It took them a long time to perform their measurements and it was a risky enterprise. At the beginning of the revolution, France entered in a war with Spain. Just try to imagine how difficult it is to explain that you are trying to define a universal unit of length when you are arrested at the top of a mountain with very precise optical instruments allowing you to follow all the movements of the troops in the surrounding. Both Delambre and Méchain were trying to reach the utmost precision in their measurements and an important part of the delay came from the fact that this reached an obsessive level in the case of Méchain. In fact when he measured the latitude of Barcelone he did it from two different close by locations, but found contradictory results which were discordant by 3.5 seconds of arc. Pressed to give his result he chose to hide this discrepancy just to “save the face” which is the wrong attitude for a Scientist. Chased from Spain by the war with France he had no second chance to understand the origin of the discrepancy and had to fiddle a little bit with his results to present them to the International Commission which met in Paris in 1799 to collect the results of Delambre and Méchain and compute the “mètre” from them. Since he was an honest man obsessed by precision, the above discrepancy kept haunting him and he obtained from the Academy to lead another expedition a few years later to triangulate further into Spain. He went and died from malaria in Valencia. After his death, his notebooks were analysed by Delambre who found the discrepancy in the measurements of the latitude of Barcelone but could not explain it. The explanation was found 25 years after the death of Méchain by a young astronomer by the name of Nicollet, who was a student of Laplace. Méchain had done in both of the sites he had chosen in Barcelone (Mont Jouy and Fontana del Oro) a number of measurements of latitude using several reference stars. Then he had simply taken the average of his measurements in each place. Méchain knew very well that refraction distorts the path of light rays which creates an uncertainty when you use reference stars that are close to the horizon. But he considered that the average result would wipe out this problem. What Nicollet did was to ponder the average to eliminate the uncertainty created by refraction and, using the measurements of Méchain, he obtained a remarkable agreement (0.4 seconds i.e. a few meters) between the latitudes measured from Mont Jouy and Fontana del Oro. In
other words Méchain had made no mistake in his measurements and could have understood
by pure thought what was wrong in his computation. I recommend the book of Ken Adler\footnote{The Measure of All Things: The Seven-Year Odyssey that Transformed the World by Ken Adler, NY: Free Pr., 2002} for a nice account of the full story of the two expeditions. In any case in the meantime
the International commission had taken the results from the two expeditions and computed
the length of the ten millionth part of the quarter of the meridian using them. Moreover a
concrete platinum bar with approximately that length was then realized and was taken as
the definition of the unit of length in the metric system. With this unit the actual length
of the quarter of meridian turns out to be 10002290 rather than the aimed for 10000000 but
this is no longer relevant. In fact in 1889 the reference became another specific metal bar
(of platinum and iridium) which was deposited near Paris in the pavillon de Breteuil. This
definition held until 1960.

Already in 1927, at the seventh conference on the metric system, in order to take into account
the inevitable natural variations of the concrete “mètre-étalon”, the idea emerged to compare
it with a reference wave length (the red line of Cadmium). Around 1960 the reference to the
“mètre-étalon” was finally abandoned and a new definition of the “mètre” was adopted as

\textbf{Figure 7.} The method of triangulation
The first accurate atomic clock was built in 1955. This led to the internationally agreed definition of the second being based on atomic time. The second is currently defined as the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the caesium atom.

1650763.73 times the wave length of the radiation corresponding to the transition between the levels 2p10 and 5d5 of the Krypton 86Kr. In 1967 the second was defined as the duration of 9192631770 periods of the radiation corresponding to the transition between the two hyperfine levels of Caesium-133. Finally in 1983 the “mètre” was defined as the distance traveled by light in 1/299792458 second. In fact the speed of light is just a conversion factor and to define the “mètre” one gives it the specific value of

\[ c = 299792458 \, \text{m/s} \]

In other words the “mètre” is defined as a certain fraction \( \frac{9192631770}{299792458} \sim 30.6633... \) of the wave length of the radiation coming from the transition between the above hyperfine levels of the Caesium atom.

The advantages of the new standard of length are many. First by not being tied up with any specific location it is in fact available anywhere where Caesium is. The choice of Caesium as opposed to Helium or Hydrogen which are much more common in the universe is of course still debatable, and it is quite possible that a new standard will soon be adopted involving spectral lines of Hydrogen instead of Caesium.

While it would be difficult to communicate our standard of length with other extra terrestrial civilizations if they had to make measurements of the earth (such as its size) the spectral
definition can easily be encoded in a probe and sent out. In fact spectral patterns (Figure 10) provide a perfect “signature” of chemicals, and a universal information available anywhere where these chemicals can be found, so that the wave length of a specific line is a perfectly acceptable unit.

5. Spectral Geometry

It is natural to wonder whether one can adapt the basic paradigm of geometry to the new standard of length, described in the previous section, which is of spectral nature. The Riemannian paradigm is based on the Taylor expansion in local coordinates $x^\mu$ of the square of the line element\[13\] in the form

$$ds^2 = g_{\mu \nu} dx^\mu dx^\nu$$

\[13\]Note that one should not confuse the “line element” $ds$ with the unit of length. In the classical framework, the latter allows one to give a numerical value to the distance between nearby points in the form (3). Multiplying the unit of length by a scalar $\lambda$ one divides the line element $ds$ by $\lambda$ since $ds$ is measured by its ratio with the unit of length.
and the measurement of the distance between two points is given by the geodesic formula

\[ d(A, B) = \operatorname{Inf} \int_{\gamma} \sqrt{g_{\mu \nu} \, dx^\mu \, dx^\nu} \]

where the infimum is taken over all paths \( \gamma \) from \( A \) to \( B \). A striking feature of this formula (4) for measuring distances is that it involves taking a square root. It is often true that “taking a square root” in a brutal manner as in (4) is hiding a deeper level of understanding. In fact this issue of taking the square root led Dirac to his famous analogue of the Schrödinger equation for the electron and the theoretical discovery of the positron.

Dirac was looking for a relativistic invariant form of the Schrödinger equation. One basic property of that equation is that it is of first order in the time variable. The Klein-Gordon equation which is the relativistic form of the Laplace equation, is relativistic invariant but is of second order in time. Dirac found a way to take the square root of the Klein-Gordon operator using Clifford algebra. In fact (as pointed out to me by Atiyah) Hamilton had already written the magic expression using his quaternions. When I was in St. Petersburg for Euler’s 300’th, I noticed that Euler could almost be credited for quaternions since he had explicitly written their multiplication rule in order to show that the product of two sums of 4 squares is a sum of 4 squares.

So what is the relation between Dirac’s square root of the Laplacian and the above issue of taking the square root in the formula for the distance \( d(A, B) \). The point is that one can use Dirac’s solution and rewrite the same distance function (4) in the following manner,

\[ d(A, B) = \operatorname{Sup} \{|f(A) - f(B)| ; f \in A, \|[D, f]\| \leq 1\} \]

where one uses the quantum set-up of [2]. Here \([D, f] = Df - fD\) of the two operators and \(\|[D, f]\|\) is its size (as explained in a footnote above). The space \( X \) is encoded through the algebra \( A \) of functions on \( X \) and, as in [2], this algebra is concretely represented as operators in Hilbert space \( \mathcal{H} \) which is here the Hilbert space of square integrable spinors. The ingredient that allows to measure distances is the operator \( D \), which is the Dirac operator in the above case of ordinary Riemannian geometry. The new formula (5) gives the same result as the geodesic distance formula (4) but it is of a quite
different nature. Indeed, instead of drawing a path from $A$ to $B$ in our space, \textit{i.e.} a map from the interval $[0,1]$ to $X$, we use maps from $X$ to the real line. In general for a given space $X$ it is not possible to join any two points by a continuous path, and thus the formula (4) is limited in its applications to spaces which are \textit{arcwise connected}. On the opposite by a theorem of topology any compact space $X$ admits plenty of continuous functions and the corresponding algebra of continuous functions can be concretely represented in Hilbert space. Thus the new formula (5) can serve as the basis for the measurement of distances and hence of the geometry of the space, provided we fix the operator $D$. Now what is the intuitive meaning of $D$? Note that formula (5) is based on the lack of commutativity between $D$ and the coordinates $f$ on our space. Thus there should be a tension that prevents $D$ from commuting with the coordinates. This tension is provided by the following key hypothesis \textquotedblleft the inverse of $D$ is an infinitesimal\textquotedblright. Indeed we saw in §2 that variables with continuous range cannot commute with infinitesimals, which gives the needed tension. But there is more, because of the following equation which gives to the inverse of $D$ the heuristic meaning of the \textit{line element}:

$$ds = D^{-1}.$$ 

Thus one can think of a geometry as a concrete Hilbert space representation not only of the algebra of coordinates on the space $X$ we are interested in, but also of its infinitesimal

\footnote{Note that $D$ has the dimension of the inverse of a length.}
line element $ds$. In the usual Riemannian case this representation is moreover irreducible. Thus in many ways this is analogous to thinking of a particle as Wigner taught us, i.e. as an irreducible representation (of the Poincaré group).

One immediate advantage of this point of view on geometry is that it uses directly the algebra of coordinates on the space $X$ one is interested in, rather than its set theoretic nature. In particular it no longer demands that this algebra be commutative.

There are very natural geometric spaces such for instance as the space of geodesics of a Riemann surface, or more generally spaces of leaves of foliations which even though they are “sets” are better described by a noncommutative algebra. The point is simple to understand: these spaces are constructed by the process of passing to the quotient and this cannot be done at once since it is impossible to select a representative in each equivalence class. Thus the quotient is encoded by the equivalence relation itself and this generates the noncommutativity, exactly as in the case of the original discovery of Heisenberg from the Ritz-Rydberg composition law. In case it is possible to select a representative in each equivalence class (as it is for instance for the geodesics on the round sphere of Figure 9), the obtained algebra is equivalent (in a sense called “Morita equivalence” on which we shall elaborate a bit below) to a commutative one, so that nothing is lost in simplicity. But one can now comprehend algebraically spaces of geometric nature which would seem untractable otherwise, a simple example being the space of geodesics on a torus, or even simpler the subspace obtained by restricting to those geodesics with a given slope. One obtains in this way the noncommutative torus, hard to “visualize” but whose subtleties are perfectly accessible through algebra and analysis.

In noncommutative geometry a space $X$ is described by the corresponding algebra $\mathcal{A}$ of coordinates which is now no longer assumed to be commutative i.e. by an involutive algebra $\mathcal{A}$ concretely represented as operators in Hilbert space $\mathcal{H}$ and the line element $ds$ which is an infinitesimal, concretely represented in the same Hilbert space $\mathcal{H}$. Equivalently a noncommutative geometry is given by a spectral triple $(\mathcal{A}, \mathcal{H}, D)$, where $D$ is the inverse of $ds$. Thus in noncommutative geometry the basic classical formula (4) is replaced by (5) where $D$ is the inverse of the line element $ds$.

The new paradigm of spectral triples passes a number of tests to qualify as a replacement of Riemannian geometry in the noncommutative world:

- It contains the Riemannian paradigm as a special case.
- It does not require the commutativity of coordinates.
- It covers the spaces of leaves of foliations.
- It covers spaces of fractal, complex, or infinite dimension.
- It applies to the analogue of symmetry groups (compact quantum groups).
- It provides a way of expressing the full standard model coupled to Einstein gravity as pure gravity on a modified spacetime geometry.
- It allows for quantum corrections to the geometry.

The reason why the quantum corrections can be taken into account is that the physics meaning of the line element $ds$ is as the propagator for Fermions, which receives quantum corrections in the form of “dressing”. The fact that the paradigm of spectral triples is compatible with compact quantum groups is the result of a long saga.
The traditional notions of geometry all have natural analogues in the spectral framework. We refer to [9] for more details. Some notions become more elaborate such as that of dimension. The dimension of a noncommutative geometry is not a number but a spectrum, the \textit{dimension spectrum} which is the subset of the complex plane at which the spectral functions have singularities.

Noncommutativity also brings new features which have no counterpart in the commutative world. In fact we already saw this in the case of the inner automorphisms which are trivial in the commutative case but correspond to the internal symmetries in physics in the noncommutative case. A similar phenomenon occurs also at the level of the metric \textit{i.e.} of the operator $D$. To understand what happens one needs to realize first that some noncommutative algebras can be intimately related to each other without being the same. For instance an algebra $\mathcal{A}$ shares most of its properties with the algebra $M_n(\mathcal{A})$ of $n \times n$ matrices over $\mathcal{A}$. Of course if $\mathcal{A}$ is commutative this is no longer the case for $M_n(\mathcal{A})$, but the theory of “vector spaces\(^{15}\) over $\mathcal{A}$” is unchanged when one replaces $\mathcal{A}$ by $M_n(\mathcal{A})$ or more generally by the algebra $\mathcal{B}$ of endomorphisms of such a vector space over $\mathcal{A}$. This relation between noncommutative algebras is called Morita equivalence and gives an equivalence relation between algebras. At the intuitive level two Morita equivalent algebras describe the same noncommutative space. Given a spectral geometry $(\mathcal{A}, \mathcal{H}, D)$ one would like to transport the geometry to a Morita equivalent algebra $\mathcal{B}$ as above. This is easily done for the Hilbert space, but when one tries to transport the metric, \textit{i.e.} the operator $D$ one finds that there is a choice involved: that of a connection. Any algebra is Morita equivalent to itself and the above ambiguity generates a bunch of metrics $D_A$ which are “internally related” to the given one $D$. In this way one gets the inner deformations of the spectral geometry, which will account for the gauge bosons in the physics context.

So much for the “metric” aspect but what about the choice of sign which is often involved when one is taking a square root and hence should play a role here since our line element $ds$ is based on Dirac’s square root for the Laplacian. Now this choice of sign, which amounts in the case of ordinary geometry to the choice of a spin structure which is the substitute for the choice of an orientation, turns out to be deeply related to the very notion of “manifold”. While we have a good recipe for constructing manifolds by gluing charts together, it is far less obvious to understand in a conceptual manner the global properties which characterize the spaces underlying manifolds. For instance the same homotopy type can underly quite different manifolds which are not homeomorphic to each other and are distinguished by the fundamental invariant given by the Pontrjagin classes. In first approximation, neglecting subtleties coming from the role of the fundamental group and also of the special properties of dimension 4, the choice of a manifold in a given homotopy type is only possible if a strong form of Poincaré duality holds (cf. [21]). The usual Poincaré duality uses the fundamental class in ordinary homology to yield an isomorphism between homology and cohomology. The more refined form of Poincaré duality that is involved in the choice of a manifold in a given homotopy type involves a finer homology theory called $KO$-homology. The operator theoretic realization of cycles in $KO$-homology was pioneered by Atiyah and Singer as a byproduct of their Index Theorem and has reached a definitive form in the work of Kasparov. The theory

\(^{15}\)In technical terms one deals with finite projective modules over $\mathcal{A}$. 
is periodic with period 8 and one may wonder how a number modulo 8 can appear in its formulation. This manifests itself in the form of a table of signs

<table>
<thead>
<tr>
<th>n</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ε</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ε'</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ε''</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

which governs the commutation relations between two simple decorations of spectral triples (with the second only existing in the even dimensional case) so that they yield a $KO$-homology class:

1. An antilinear isometry $J$ of $\mathcal{H}$ with $J^2 = \epsilon$ and $J D = \epsilon' D J$.
2. A $\mathbb{Z}/2$-grading $\gamma$ of $\mathcal{H}$, such that $J \gamma = \epsilon'' \gamma J$.

The three signs $(\epsilon, \epsilon', \epsilon'') \in \{\pm 1\}^3$ detect the dimension modulo 8 of the cycle in $KO$-homology, as ruled by the above table.

These decorations have a deep meaning both from the mathematical point of view – where the main underlying idea is that of a manifold – as well as from the physics point of view. In the physics terminology the operator $J$ is the charge conjugation operator, and the grading $\gamma$ is the chirality.

The compatibilities of these simple decorations with the algebra $\mathcal{A}$ are the following; the antilinear involution $J$ fulfills the “order zero” condition:

$$[a, J b^* J^{-1}] = 0, \quad \forall a, b \in \mathcal{A},$$

and the $\mathbb{Z}/2$-grading $\gamma$ fulfills $\gamma A \gamma^{-1} = A$, which gives a $\mathbb{Z}/2$-grading of the algebra $\mathcal{A}$.

The mathematical origin of the commutation relation (6) is in the work of the Japanese mathematician M. Tomita who proved that, given an involutive algebra $\mathcal{A}$ of operators in $\mathcal{H}$, one obtains an antilinear involution $J$ fulfilling (6) by taking the polar decomposition of the operator

$$a \xi \mapsto a^* \xi, \quad \forall a \in \mathcal{A}$$

under the key hypothesis that the vector $\xi$ is separating for $\mathcal{A}$ i.e. that $\mathcal{A}' \xi = \mathcal{H}$ where $\mathcal{A}'$ is the commutant of $\mathcal{A}$.

Thus, to summarize, a real spectral triple (i.e. a spectral triple $(\mathcal{A}, \mathcal{H}, D)$ together with the decorations $J$ and $\gamma$) encodes both the metric and the fundamental class in $KO$-homology for a noncommutative space. A key condition that plays a role in characterizing the classical case is the following “order one” condition relating the operator $D$ with the algebra:

$$[[D, a], b^0] = 0 \quad \forall a, b \in \mathcal{A}.$$  

where we use the notation $b^0 = J b^* J^{-1}$ for any $b \in \mathcal{A}$.

6. The “raison d’être” of the finite space $F$

Poincaré and Einstein showed how to infer the correct flat spacetime geometry (known as Minkowski space geometry) underlying special relativity from experimental evidence associated to Maxwell’s theory of electromagnetism. In fact, the Maxwell equations are intrinsically
relativistic. This flat geometry is then extended to the curved Lorentzian manifolds of general relativity. From the particle physics viewpoint, the Lagrangian of electromagnetism is however just a very small part of the full Standard Model Lagrangian (cf. Figure 3). Thus, it is natural to wonder whether the transition from the Lagrangian of electrodynamics to the Standard Model can be understood as a further refinement of the geometry of spacetime, rather than the introduction of a zoo of new particles.

The spectral formalism of noncommutative geometry, explained above, makes it possible to consider spaces which are more general than ordinary manifolds. This gives us more freedom to obtain a suitable geometric setting that accounts for the additional terms in the Lagrangian.

The idea to obtain the spectral geometry \((\mathcal{A}, \mathcal{H}, D)\) is that:

- The algebra \(\mathcal{A}\) is dictated by the comparison of its group of inner automorphisms with the internal symmetries.
- The Hilbert space \(\mathcal{H}\) is the Hilbert space of Euclidean fermions.
- The line element \(ds\) is the propagator for Euclidean fermions.

Using this idea together with the spectral action principle (\(\mathcal{L}\)), which will be explained in the next section, allowed to determine a very specific finite noncommutative geometry \(F\) such that pure gravity (in the form of the spectral action) on the product \(M \times F\), with the metric given by the inner fluctuations of the product metric, delivers the Standard Model coupled to gravity.

Thus in essence what happens is that the scrutiny of spacetime at very small scales (of the order of \(10^{-16}\) cm) reveals a fine structure which replaces an ordinary point in the continuum by the finite geometry \(F\).

In the above approach this finite geometry was taken from the phenomenology i.e. put by hand to obtain the Standard Model Lagrangian using the spectral action. The algebra \(\mathcal{A}_F\), the Hilbert space \(\mathcal{H}_F\) and the operator \(D_F\) for the finite geometry \(F\) were all taken from the experimental data. The algebra comes from the gauge group, the Hilbert space has
as a basis the list of elementary fermions and the operator is the Yukawa coupling matrix. This worked fine for the minimal Standard Model, but there was a problem of doubling the number of Fermions, and also the Kamiokande experiments on solar neutrinos showed around 1998 that, because of neutrino oscillations, one needed a modification of the Standard Model incorporating in the leptonic sector of the model the same type of mixing matrix already present in the quark sector. One further needed to incorporate a subtle mechanism, called the see-saw mechanism, that could explain why the observed masses of the neutrinos would be so small. At first my reaction to this modification of the Standard Model was that it would certainly not fit with the noncommutative geometry framework and hence that the previous agreement with noncommutative geometry was a mere coincidence. I kept good relations with my thesis adviser Jacques Dixmier and he kept asking me to give it a try, in spite of my “a priori” pessimistic view. He was right and after about 8 years I realized (cf. [12]) that the only needed change (besides incorporating a right handed neutrino per generation) was to make a very simple change of sign in the grading for the anti-particle sector of the model (this was also done independently in [1]). This not only delivered naturally the neutrino mixing, but also gave the see-saw mechanism and settled the above Fermion doubling problem. The main new feature that emerges is that when looking at the above table of signs giving the KO-dimension, one finds that the finite noncommutative geometry $F$ is now of dimension 6 modulo 8! Of course the space $F$ being finite, its metric dimension is 0 and its inverse line-element is bounded. In fact this is not the first time that spaces of this nature—i.e. whose metric dimension is not the same as the KO-dimension—appear in noncommutative geometry and this phenomenon had already appeared for quantum groups.

Besides yielding the Standard Model with neutrino mixing and making testable predictions (as we shall see in §8), this allowed one to hope that, instead of taking the finite geometry $F$ from experiment, one should in fact be able to derive it from first principles. The main intrinsic reason for crossing by a finite geometry $F$ has to do with the value of the dimension of spacetime modulo 8. We would like this KO-dimension to be 2 modulo 8 (or equivalently 10) to define the Fermionic action, since this eliminates the doubling of fermions in the Euclidean framework. In other words the need for crossing by $F$ is to shift the KO-dimension from 4 to 2 (modulo 8).

This suggested to classify the simplest possibilities for the finite geometry $F$ of KO-dimension 6 (modulo 8) with the hope that the finite geometry $F$ corresponding to the Standard Model would be one of the simplest and most natural ones. This was finally done recently in our joint work with A. Chamseddine ([8]).

From the mathematical standpoint our road to $F$ is through the following steps

1. We classify the irreducible triplets $(A, H, J)$.
2. We study the $\mathbb{Z}/2$-gradings $\gamma$ on $H$.
3. We classify the subalgebras $A_F \subset A$ which allow for an operator $D$ that does not commute with the center of $A$ but fulfills the “order one” condition:

$$[[D, a], b^0] = 0 \quad \forall a, b \in A_F.$$

---

16 joint work [2] with Ali Chamseddine

17 because this allows one to use the antisymmetric bilinear form $\langle J\xi, D\eta \rangle$ (for $\xi, \eta \in H, \gamma\xi = \xi, \gamma\eta = \eta$)
The classification in the first step shows that the solutions fall in two classes, in the first the dimension \( n \) of the Hilbert space \( \mathcal{H} \) is a square: \( n = k^2 \), in the second case it is of the form \( n = 2k^2 \). In the first case the solution is given by a real form of the algebra \( M_k(\mathbb{C}) \) of \( k \times k \) complex matrices. The representation is given by the action by left multiplication on \( \mathcal{H} = M_k(\mathbb{C}) \), and the isometry \( J \) is given by \( x \in M_k(\mathbb{C}) \mapsto J(x) = x^* \). In the second case the algebra is a real form of the sum \( M_k(\mathbb{C}) \oplus M_k(\mathbb{C}) \) of two copies of \( M_k(\mathbb{C}) \) and while the action is still given by left multiplication on \( \mathcal{H} = M_k(\mathbb{C}) \oplus M_k(\mathbb{C}) \), the operator \( J \) is given by \( J(x, y) = (y^*, x^*) \).

The study \((2)\) of the \( \mathbb{Z}/2 \)-gradings shows that the commutation relation \( J\gamma = -\gamma J \) excludes the first case. We are thus left only with the second case and we obtain among the very few choices of lowest dimension the case \( \mathcal{A} = M_2(\mathbb{H}) \oplus M_4(\mathbb{C}) \) where \( \mathbb{H} \) is the skew field of quaternions. At a more invariant level the Hilbert space is then of the form \( \mathcal{H} = \text{Hom}_{\mathbb{C}}(V, W) \oplus \text{Hom}_{\mathbb{C}}(W, V) \) where \( V \) is a 4-dimensional complex vector space, and \( W \) a two dimensional graded right vector space over \( \mathbb{H} \). The left action of \( \mathcal{A} = \text{End}_{\mathbb{H}}(W) \oplus \text{End}_{\mathbb{C}}(V) \) is then clear and its grading as well as the grading of \( \mathcal{H} \) come from the grading of \( W \).

Our main result then is that there exists up to isomorphism a unique involutive subalgebra of maximal dimension \( \mathcal{A}_F \) of \( \mathcal{A}^{ev} \), the even part\(^{18}\) of the algebra \( \mathcal{A} \), which solves (3). This involutive algebra \( \mathcal{A}_F \) is isomorphic to \( \mathbb{C} \oplus \mathbb{H} \oplus M_3(\mathbb{C}) \) and together with its representation in \( (\mathcal{H}, J, \gamma) \) gives the noncommutative geometry \( F \) which we used in \([6]\) to recover the Standard Model coupled to gravity using the spectral action which we now describe.

7. Observables in gravity and the spectral action

The missing ingredient, in the above description of the Standard Model coupled to gravity, is provided by a simple action principle—the spectral action principle \([2], [3], [4], [5]\) — that has the geometric meaning of “pure gravity” and delivers the action functional \((1)\) when evaluated on \( M \times F \). To this action principle we want to apply the criterion of \textit{simplicity} rather than that of \textit{beauty} given the relative nature of the latter. Thus we imagine trying to explain this action principle to a Neanderthal man. The spectral action principle, described below, passes the “Neanderthal test”, since it amounts to counting spectral lines.

The starting point at the conceptual level is the discussion of observables in gravity. By the principle of gauge invariance the only quantities which have a chance to be observable in gravity are those which are invariant under the group of diffeomorphisms of the spacetime \( M \). Assuming first that we deal with a classical manifold (and Wick rotate to euclidean signature for simplicity), one can form a number of such invariants (under suitable convergence conditions) as the integrals of the form

\[ \int_M F(K) \sqrt{g} \, d^4x \]

where \( F(K) \) is a scalar invariant function\(^{19}\) of the Riemann curvature \( K \). There are\(^{20}\) other more complicated examples of such invariants, where those of the form \((8)\) appear as the \textit{single integral} observables \textit{i.e.} those which add up when evaluated on the direct sum of geometric

---

\(^{18}\)One restricts to the even part to obtain an ungraded algebra.

\(^{19}\)The scalar curvature is one example of such a function but there are many others

spaces. Now while in theory a quantity like (8) is observable it is almost impossible to evaluate since it involves the knowledge of the entire spacetime and is in that way highly non localized. On the other hand, spectral datas are available in localized form anywhere, and are (asymptotically) of the form (8) when they are of the additive form

\[ \text{Trace} \left( f(D/\Lambda) \right), \]

where \( D \) is the Dirac operator and \( f \) is a positive even function of the real variable while the parameter \( \Lambda \) fixes the mass scale. The spectral action principle asserts that the fundamental action functional \( S \) that allows to compare different geometric spaces at the classical level and is used in the functional integration to go to the quantum level, is itself of the form (9). The detailed form of the function \( f \) is largely irrelevant since the spectral action (9) can be expanded in decreasing powers of the scale \( \Lambda \) and the function \( f \) only appears through the scalars

\[ f_k = \int_0^\infty f(v) v^{k-1} dv. \]

As explained above the gauge potentials make good sense in the framework of noncommutative geometry and come from the inner fluctuations of the metric.

Let \( M \) be a Riemannian spin 4-manifold and \( F \) the finite noncommutative geometry of \( KO \)-dimension 6 described above. Let \( M \times F \) be endowed with the product metric. Then by [6]

1. The unimodular subgroup of the unitary group acting by the adjoint representation \( \text{Ad}(u) \) in \( \mathcal{H} \) is the group of gauge transformations of SM.
2. The unimodular inner fluctuations of the metric give the gauge bosons of SM.
3. The full standard model (with neutrino mixing and seesaw mechanism) minimally coupled to Einstein gravity is given in Euclidean form by the action functional

\[ S = \text{Tr}(f(D_A/\Lambda)) + \frac{1}{2} \langle J \tilde{\xi}, D_A \tilde{\xi} \rangle, \quad \tilde{\xi} \in \mathcal{H}_d^+, \]

where \( D_A \) is the Dirac operator with the unimodular inner fluctuations.

The change of variables from the standard model to the spectral model is summarized in Table 1. We refer to [6] for the notations.

8. Predictions

The above spectral model can be used to make predictions assuming the “big desert” (absence of new physics up to unification scale) together with the validity of the spectral action as an effective action at unification scale. While the big-desert hypothesis is totally improbable, a rough agreement with experiment would be a good indication for the spectral model.

When a physical theory is described at the classical level by an action principle i.e. by minimizing an action functional \( S \) on the configurations, there is a heuristic prescription due to R. Feynman in order to go from the classical to the quantum. This prescription affects each classical field configuration with the probability amplitude:
<table>
<thead>
<tr>
<th>Standard Model</th>
<th>notation</th>
<th>notation</th>
<th>Spectral Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higgs Boson</td>
<td>$\varphi = (\frac{2M}{g} + H - i\phi^0, -i\sqrt{2}\phi^+)$</td>
<td>$H = \frac{1}{\sqrt{2}} \sqrt{g}(1 + \psi)$</td>
<td>Inner metric(^{(0,1)})</td>
</tr>
<tr>
<td>Gauge bosons</td>
<td>$A_\mu, Z_\mu, W^\pm_\mu, g^a_\mu$</td>
<td>$(B, W, V)$</td>
<td>Inner metric(^{(1,0)})</td>
</tr>
<tr>
<td>Fermion masses</td>
<td>$m_u, m_\nu$</td>
<td>$Y_{(13)} = \delta_{(13)}$, $Y_{(11)} = \delta_{(11)}$</td>
<td>Dirac(^{(0,1)}) in ↑</td>
</tr>
<tr>
<td>CKM matrix</td>
<td>$C^c_X, m_d$</td>
<td>$Y_{(13)} = C \delta_{3,1} C^\dagger$</td>
<td>Dirac(^{(0,1)}) in ↓</td>
</tr>
<tr>
<td>Lepton mixing</td>
<td>$U^{lep}_{\lambda \kappa}, m_e$</td>
<td>$Y_{(11)} = U^{lep} \delta_{(11)} U^{lep\dagger}$</td>
<td>Dirac(^{(0,1)}) in ↓</td>
</tr>
<tr>
<td>Majorana mass</td>
<td>$M_R$</td>
<td>$Y_R$</td>
<td>Dirac(^{(0,1)}) on $E_R \oplus J_F E_R$</td>
</tr>
<tr>
<td>mass matrix</td>
<td>$g_1 = g \tan(\theta_w)$, $g_2 = g$, $g_3 = g_s$</td>
<td>$g_3^2 = g_2^2 = \frac{2}{3} g_1^2$</td>
<td>Fixed at unification</td>
</tr>
<tr>
<td>Gauge couplings</td>
<td>$\lambda_0 = g^2 \frac{h}{v^2}$</td>
<td>$\lambda_0 = g^2 \frac{h}{v^2}$</td>
<td>Fixed at unification</td>
</tr>
<tr>
<td>Higgs scattering parameter</td>
<td>$\frac{1}{8} g^2 \alpha_h, \alpha_h = \frac{m^2}{4M^2}$</td>
<td>$\lambda_0 = g^2 \frac{h}{v^2}$</td>
<td>Fixed at unification</td>
</tr>
<tr>
<td>Tadpole constant</td>
<td>$\beta_h, (-\alpha_h M^2 + \frac{\beta_h}{2})</td>
<td>\varphi</td>
<td>^2$</td>
</tr>
<tr>
<td>Graviton</td>
<td>$g_{\mu \nu}$</td>
<td>$\tilde{\theta}_M$</td>
<td>Dirac(^{(1,0)})</td>
</tr>
</tbody>
</table>

**Table 1.** Conversion from Spectral Action to Standard Model

$$\mathcal{E}^i \frac{S}{\hbar}$$

This prescription is only heuristic for a number of reasons, one being that the overall sum over all configurations is an oscillatory integral whose convergence is hard to control. This situation improves if one works in the Euclidean formulation. This means that our configurations are Euclidean and in the functional integral the weight of such a configuration is now given by
After passing to the Euclidean formulation, one can start computing the Feynman integral applying perturbative techniques, but one quickly meets a basic problem due to the omnipresence of divergent integrals in evaluating the contribution of simple processes (called Feynman graphs) such as the emission and absorption of a photon by an electron shown in Figure 12. This problem was already present before the Feynman path integral formulation, when trying to compute the higher order terms in Dirac’s theory of absorption and emission of radiation. Physicists discovered around 1947 a procedure, called renormalization, to handle this problem. The physics idea is to make the distinction between the so-called bare parameters which enter in the mathematical formula for the action $S$ and the observed parameters such as masses, charges etc... This distinction goes back to the work of Green in hydrodynamics in the nineteenth century. It is easy to explain in a simple example: the motion of a ping-pong ball inside water (Figure 13). What Green found is that Newton’s law $F = ma$ remains valid, but the mass $m$ is no longer the bare mass $m_0$ that would be obtained by weighing the ping-pong ball with a scale outside the water, but a modified mass

$$m = m_0 + \frac{1}{2}M$$

where $M$ is the mass of the same volume of water. In other words, due to the presence of the surrounding field of water the inertial mass of the ball is increased, as if it were half full of water! This “renormalization” of mass has direct measurable effects even on the initial
acceleration of the ping-pong ball shooting back to the surface (it is about 7-times smaller than what one would compute using the Archimedean law).

In order to take care of the divergent integrals that are omnipresent in the perturbative calculations of quantum field theory, one introduces an energy scale $\Lambda$ (called a “cutoff” scale) and one only restricts the integration variable to values of the energy which are smaller than $\Lambda$ (this is the “cutoff”). One then cleverly lets the “constants” which enter in the formula for the action $S$ depend on $\Lambda$ so that the infinities disappear (in the computations of observable quantities) when one removes the cutoff, i.e. when $\Lambda$ goes to infinity. This magic way of sweeping the difficulty under the rug works for theories which are “renormalizable” and gives in electrodynamics incredibly precise predictions which, for the anomalous moment of the electron, are in perfect agreement with the observed value. The price one pays is that this variability of the constants (they now depend on $\Lambda$) introduces a fundamental ambiguity in the renormalization process.

The corresponding symmetry group is called the renormalization group, and in a recent work with M. Marcolli we exhibited an incarnation of this group as a universal symmetry group—
called the “cosmic Galois group” following P. Cartier—of all renormalizable theories. As the name suggests it is deeply related to the ideas of Galois which he briefly sketched in his last letter:

"Tu sais, mon cher Auguste, que ces sujets ne sont pas les seuls que j'aie explorés. Mes principales méditations depuis quelque temps étaient dirigées sur l'application à l'analyse transcendant de la théorie de l'ambiguïté. Il s'agissait de voir a priori dans une relation entre des quantités ou fonctions transcendantes quels échanges on pouvait faire, quelles quantités on pouvait substituer aux quantités données sans que la relation puisse cesser d'avoir lieu. Cela fait reconnaître tout de suite l'impossibilité de beaucoup d'expressions que l'on pourrait chercher. Mais je n'ai pas le temps et mes idées ne sont pas encore assez développées sur ce terrain qui est immense”

It might seem that the presence of the above ambiguity precludes the possibility to predict the values of these physical constants, which are in fact not constant but depend upon the energy scale $\Lambda$. In fact the renormalization group gives differential equations which govern their dependence upon $\Lambda$. The intuitive idea behind this equation is that one can move down, i.e. lower the value of $\Lambda$ to $\Lambda - d\Lambda$ by integrating over the modes of vibrations which have their frequency in the given interval. For the three coupling constants $g_i$ (or rather their square $\alpha_i$) which govern the three forces (excluding gravity) of the Standard Model, their dependence upon the scale is shown in Figure 15 and shows that they are quite different at low scale, they become comparable at scales of the order of $10^{15}$ GeV. This suggested long ago the idea that physics might become simpler and “unified” at scales (called unification scales) of that order.

We can now describe the predictions obtained by comparing the spectral model with the standard model coupled to gravity. The status of “predictions” in the above spectral model is based on two hypothesis:

1. The model holds at unification scale
2. One neglects the new physics up to unification scale

The value of that scale is the above unification scale since the spectral action delivers the same equality $g_2^2 = g_3^2 = \frac{5}{3} g_1^2$ which is common to all “Grand-Unified” theories. It gives more precisely the following unification of the three gauge couplings:

$$\frac{g_2^2 f_0}{2\pi^2} = \frac{1}{4}, \quad g_3^2 = g_2^2 = \frac{5}{3} g_1^2.$$  

Here $f_0 = f(0)$ is the value of the test function $f$ at 0.

The second feature which is predicted by the spectral model is that one has a see-saw mechanism for neutrino masses with large $M_R \sim \Lambda$.

The third prediction that one gets by making the conversion from the spectral model to the standard model is that the mass matrices satisfy the following constraint at unification scale:

$$\sum_\sigma \left( m_{\nu\sigma}^2 \right)^2 + \left( m_{e\sigma}^2 \right)^2 + 3 \left( m_{u\sigma}^2 \right)^2 + 3 \left( m_{d\sigma}^2 \right)^2 = 8 M_W^2.$$
In fact it is better to formulate this relation using the following quadratic form in the Yukawa couplings:

\[ Y_2 = \sum_{\sigma} (y_{\sigma}^\nu)^2 + (y_{\sigma}^e)^2 + 3(y_{\sigma}^u)^2 + 3(y_{\sigma}^d)^2 \]

so that the above prediction means that

\[ Y_2(S) = 4g^2. \]

This, using the renormalization group (cf. Figure 16) to compute the effective value at our scale, yields a value of the top mass which is 1.04 times the observed value when neglecting the Yukawa couplings of the bottom quarks etc...and is hence compatible with experiment.

Another prediction obtained from the conversion table is the value of the Higgs scattering parameter at unification:

\[ \tilde{\lambda}(\Lambda) = g_3^2 \frac{b}{a^2}. \]

The numerical solution to the renormalization group equations (cf. Figure 17) with the boundary value \( \lambda_0 = 0.356 \) at \( \Lambda = 10^{17} \) GeV gives \( \lambda(M_Z) \sim 0.241 \) and a Higgs mass of the order of 168 GeV.

Finally since our theory unifies the Standard Model with gravity it also predicts the value of the Newton constant at unification scale. Here one needs to be really careful since while renormalization works remarkably well for the quantization of the classical fields involved in the standard model, this latter perturbative technique fails when one tries to deal with the gravitational field \( g_{\mu\nu} \) using the Einstein-Hilbert action. However the spectral action delivers
additional terms like the square of the Weyl curvature, and one can then use this action as an effective action. This means that one does not pretend to have a fundamental theory \textit{i.e.} a theory that will run at any energy, but rather an action functional which is valid at a certain energy scale. Even then, the running of the Newton constant under the renormalization group is scheme dependent and not well understood. However making the simple hypothesis that the Newton constant does not change much up to a unification scale of the order of $10^{17}$ GeV one finds that it suffices to keep a ratio of order one between the moments of the test function $f$ ($f_2 \sim 5f_0$) to obtain a sensible value. We also checked that the additional terms such as the term in Weyl curvature square, do not have any observable consequence at ordinary scales (the running of these terms is known and scheme independent).

9. What about Quantum Gravity

As explained above the theory which is obtained from the spectral action applied to noncommutative geometries of the form $M \times F$ is not a fundamental theory but rather an effective theory, in that it stops making sense above the unification scale. For instance the gravitational propagator admits a tachyon pole and unitarity breaks down. It is thus natural to wonder how such a theory can emerge from a more fundamental one. In other words one

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure16.png}
\caption{The running of the top quark Yukawa coupling.}
\end{figure}
would like to get some idea of what happens at energies above the unification scale, if that has any meaning.

In our book [18] with M. Marcolli we have developed an analogy between the role of spontaneous symmetry breaking in a number theoretic framework—intimately related to the spectral realization of the zeros of the Riemann zeta function—and the symmetry breaking in the electroweak sector of the Standard Model.

It raises in particular the possibility that geometry only emerges after a suitable symmetry breaking mechanism which extends to the full gravitational sector the electroweak symmetry breaking that we discussed above. The invariance of the spectral action under the symplectic unitary group in Hilbert space is broken during this process to the compact group of isometries of a given geometry.

Thus, if one follows closely the analogy with the number theoretic system, one finds that above a certain energy (the Planck energy, say) the very idea of a spacetime disappears and the state of the global system is a mixed state of type III (in the sense that it generates a factor of type III) whose support consists of operator theoretic datas with essentially no classical geometric meaning. In the number theoretic system the role of the energy level is played by temperature, and the key notion is that of KMS state. Thus if one follows this analogy one needs to abandon the idea of the initial singularity of spacetime and replace it by the emergence of geometry, through a symmetry breaking phenomenon. In particular the idea of trying to quantize the gravitational field in a fixed background spacetime manifold by finding a renormalizable unitary quantum field theory becomes unrealistic.

**Figure 17.** The running of the quartic Higgs coupling.
ON THE FINE STRUCTURE OF SPACETIME

REFERENCES


[18] A. Connes, M. Marcolli, book


A. CONNES: COLLÈGE DE FRANCE, 3, RUE D’ULM, PARIS, F-75005 FRANCE, I.H.E.S. AND VANDERBILT UNIVERSITY
E-mail address: alain@connes.org