1

Dimensional dreams

1.1 Space: a historical view

Of all the aspects of the physical world, the mysteries regarding the nature of space and time are the most recondite. Questions related to the nature of space and time have engaged philosophers, mathematicians and physicists alike across histories and cultures. It appears that even in ancient cultures some thinkers concerned themselves with these questions but it was probably with the Greeks that an attempt at a systematic, physical understanding of space and time is discerned. This is not surprising given that the Greek interest in both geometry and mechanics was motivated by physical rather than metaphysical considerations. The famous antinomies of Zeno tell us that even pre-Socratic philosophers had devoted much thought to the understanding of space, time and motion. But somewhere in the period between Zeno and Aristotle, earlier meditations on these issues crystallised and were put down more concretely.

Between the Greek atomists and Aristotle, then, we have been provided two enduring conceptions of space which are broad enough to accomodate the whole gamut of ideas about space that have emerged over the centuries. For Aristotle, space did not have an existence independent of the material world but rather was a relational property of material bodies. Independent of matter, space had no existence: there was no conception of a vacuum. Any statement about space that is made is really a statement about a relation between different material entities. In contrast to Aristotle, and in fact preceding him, there existed another conception of space due to the atomists, prominent among whom were Leucippus and Democritus. The atoms, which were unchanging and eternal, gave rise to the variety and the change perceived in the phenomenal world by coming together in different ways as they moved about in the void. The void or empty space had an existence independent of the atoms that moved in them and was as fundamental as the atoms. These two views, relationalism (of Aristotle) and substantivalism (of the atomists), have been paradigmatic for theories of space that have been developed for over two millenia. While for the substantivalists, geometry has a significance of its own, for the relationalists it is something that can only be inferred from the properties of material bodies. A comprehensive discussion of the different philosophical conceptions of space and the historical evolution of these ideas is contained in Ref. [1].

It was Aristotle's relationalism that dominated until the time of Newton when substantivalism made a comeback. In his *Principia*, Newton took care to distinguish between absolute and relative motion, which he distinguished as true and apparent motion. In trying to avoid the conclusion about the relativity of all motion, Newton distinguished the notion of relative spaces from that of absolute space. Instead of accepting the fact that all inertial frames were equivalent, i.e. related to each other by a Galilean transformation and that all motion was relative, Newton stuck to the idea of one absolute space identified as the fixed, immovable centre of the world.

Newton's idea of absolute space was criticised by Leibniz, Huygens and Berkeley. But it was only in the latter half of the nineteenth century that Ernst Mach mounted a systematic critique of Newton's ideas. Mach was willing to accept only inertial systems and argued that absolute space was a metaphysical entity and had to be eliminated from mechanics. In the meantime, however, electromagnetic field theory had been developed by Maxwell and the prediction of electromagnetic radiation was verified by Hertz. This immediately brought up the question of the existence of a medium in which these waves could propagate. If it could be demonstrated that such a medium, called ether, existed then it would very much be the absolute frame with respect to which the motion of all bodies could be measured. The earth's motion through the ether meant that the speed of light measured in the direction and against the direction of the earth's motion would be different. The Michelson-Morley experiment set out to detect such a difference in the speed of light so as to establish the existence of ether. The famous null result of the experiment led to the formulation of Einstein's special relativity though, for a while, Lorentz persisted with a substantival interpretation of the experimental results. Apart from the other great changes that this new theory wrought, the special theory also made a radical change in the conception of space. The notion of absolute space was eliminated so it was a new form of relationalism. But, more importantly, through its critique of the notion of simultaneity of physical events, special relativity also emphasised how it was not possible to think of space and time independently and brought in the new notion of a spacetime continuum. This was truly a new extension of the notion of space by appending time to it as a fourth dimension, through the introduction of the Minkowski metric.[1] We have alluded to this well-known fact about

[1] The idea of treating time as the fourth dimension was first mentioned by d'Alembert in an *Encyclopédie* article that he published in 1751 [2] and later also by Lagrange in *Mécanique Analytique* [3].

relativity to emphasise a historical point: that spurred by empirical information coming from the Michelson-Morley experiment revisions of notions of space and time were being carried out demonstrating for the very first time that, in spite of the abstruseness of the notions of space and time, these were amenable to an experimental study much like other physical phenomena.

It is likely that this freedom in exploring space and time on par with other physical phenomena was what provided the psycho-socio impetus to develop the General Theory of Relativity. Space had already been liberated from the fetters of Euclideanism through the work of nineteenth-century mathematicians like Gauss, Lobachevski and, most of all, Riemann. It was possible now to move out of the flatland of Euclidean geometry and explore the curved spaces of non-Euclidean geometries. What was remarkable about Einstein's General Theory was that he related the curvature of spacetime to the local distribution of matter and energy and in bringing together dynamics and geometry he, once and for all, established that the study of the nature of space and time was an empirical discourse to be settled by comparing theories, however fanciful, with hard facts of experiment – geometry was not an *a priori* category.

1.2 An extra dimension

One of the properties of space that attracted considerable attention was its dimensionality. Why is space three-dimensional? As always, there were two different approaches taken in addressing this question. One approach was to somehow deduce three as the only possible value for the dimensionality of space. While many of these attempts seem to be only of historical interest and lack any scientific merit some of these arguments are still compelling. One such argument, advanced by Kant,[2] was that the inverse-square law of gravitation would materialise only in a three-dimensional world and the introduction of extra dimensions would modify Newton's law of gravitation. Of course, the argument rested on the assumption of the absolute validity of the law of gravitation at all length scales. In fact, it is amusing to note that modern-day searches for extra dimensions look for precisely such deviations from the inverse-square law at short distances. Ehrenfest [5] and several others[3] also argued that only for $d = 3$ are stable planetary orbits possible, though strictly speaking the stability arguments do not rule out $d < 3$.

The other way to address the issue of dimensionality was to ask if there is a possibility that there is an extra dimension – a fourth dimension, or possibly, even more dimensions. The first such attempt seems to have been made by Henry

[2] Kant published this in a 1746 manuscript entitled *Thoughts on the True Estimation of Living Forces*, a translation of which by J. B. Edwards and M. Schönfeld appears in a volume of Kant's writings on natural science [4].

[3] See Ref. [6] for a complete list of these papers and for the details of the arguments.

More, a neo-Platonist philosopher from Cambridge who was a contemporary of Newton. In fact, Newton's views of absolute space seem to be very strongly influenced by More's ideas. For More, like Newton, the absolute character of space was a manifestation of the omniscience of God. If that were so then space would not be just the arena for physical phenomena but for the spiritual and psychical as well. It was to do this that More thought of a fourth dimension: a geometrical dimension in which spiritual processes took place [7]. More called this new dimension *spissitude* from the Latin word for thickness. More's spissitude did not attract too many followers but he did have a dedicated fringe following all the way down to the nineteenth and twentieth centuries. The nineteenth-century astronomer from Leipzig, J. F. K. Zöllner, took More's ideas very seriously and performed several experiments to establish the existence of spissitude. Supernatural phenomena, like the sudden appearance or disappearance of objects and several other miracles were made by Zöllner the subject of this study [8]. As late as the early twentieth century, Henry More's ideas seemed to have had currency for he seemed to have considerably influenced the Portuguese poet, Fernando Pessoa.

Throughout the nineteenth century, interest in the idea of a fourth dimension persisted. At the same time as Zöllner dabbled in his experiments with the supernatural, his colleague in Leipzig, the mathematician August Möbius, demonstrated how it is possible to use a four-dimensional rotation to turn a three-dimensional object into its mirror image [9]. In 1852, the Swiss mathematician Ludwig Schläfli discovered [10] the six four-dimensional analogues of the five three-dimensional platonic solids, now known as polytopes. Schläfli's results were duplicated by Stringham in 1880 [11] but in his work, Stringham provided illustrations of projections on a plane of the four-dimensional polytopes. This work became very well-known and his depiction of the four-dimensional cube, also called the tesseract became part of the popular imagination and served as an inspiration for several works of art and literature. Much of the popularity of the tesseract was due to the writings of Charles Hinton whose book *The Fourth Dimension* [12] became immensely popular and had among its large following even the eminent American philosopher, William James.

Again, it was not until the advent of relativity that the idea of extra dimensions came into physics proper. In fact, as stated earlier, Minkowski's realisation of a geometrical interpretation of the Lorentz transformations in a four-dimensional spacetime was already a step in the direction. But Minkowski had a simpler task because he did not invent a fourth dimension but simply reinterpreted time (or ict) as a dimension and in this he was already helped by the fact that the phenomena of electricity and magnetism, now systematised in Maxwell's theory of electromagnetism, showed invariance with respect to Lorentz transformations and not the Galilean transformations. So it was, in a sense, empirical exigencies that made the way for Minkowski's realisation. The problem of theorising about a dimension beyond these four dimensions of spacetime was altogether different. Not only was there was no empirical support for such a theory but if one were to go by common-sense experience then there appeared to be no way one could accommodate any more dimensions (spatial or temporal). With the advent of General Relativity, however, came the realisation of the intimate relation of geometry and dynamics. If the gravitational interaction were to be understood as the manifestation of four-dimensional geometry then is it possible to arrive at an unified understanding of gravitation and electromagnetism by bringing in a fifth dimension? This was the question that prompted both Nordström, as early as 1914 (a year before Einstein's General Theory!) [13], and Kaluza [14], in 1919, to consider five-dimensional spacetimes. Nordström was considering a scalar theory of gravity while Kaluza, benefitting from Einstein's formulation of the General Theory, worked out a tensor theory. In order to sidestep the issue of the non-observability of the fifth dimension, both Nordström and Kaluza assumed that all derivatives of the fields with respect to the fifth-dimensional co-ordinate vanish, i.e. the fields do not depend on the fifth dimension. This condition is known as the 'cylinder' condition.[4] Kaluza then assumed a five-dimensional tensor theory of gravity but in the absence of any matter. The five-dimensional metric G_{MN} decomposes into a four-dimensional part $g_{\mu\nu}$, a vector potential A_μ and a scalar ϕ, i.e. Kaluza identified the four-dimensional part of the five-dimensional metric as the usual four-dimensional metric, which is related to gravitation, and identified the vector $G_{4\mu}$ as the electromagnetic potential A_μ. The scalar field was a bit of an embarrassment to Kaluza and he ignored it by setting it equal to unity but if included it gives rise to a Brans-Dicke-type scalar. It was noted much later [16] that the condition $\phi = 1$ leads to the unphysical condition $F_{\mu\nu}F^{\mu\nu} = 0$.

The assumption in Kaluza's theory of having no dependence of physical quantities on the fifth dimension was both drastic and unpalatable. It was after the advent of quantum mechanics that Klein came up with the suggestion [17] that one could treat this dimension on par with the other dimensions but compactify it to very small sizes, i.e. by assigning a circular topology S^1 to it and making the radius of this circle very tiny. Fields on the circle then naturally admitted of a Fourier expansion into modes labelled by n and then, as suggested in the new quantum mechanics, each mode could be assigned a momentum $|n|/R$, with R being the radius of compactification. The diminutiveness of R then makes the momentum associated with all modes $n > 0$ very large, putting them beyond the reach of observation with only the zero mode ($n = 0$) (which is independent of the fifth dimensional co-ordinate) observable. Thus by introducing a small

[4] The history of the early period of Kaluza-Klein theories has been studied in detail [15]. Einstein, to whom Kaluza communicated his paper in 1919, reacted saying that the idea was new to him. But he was certainly aware of Nordström's work who had worked out his scalar gravitation theory in a series of papers, which was considered the only competitor to Einstein's tensor theory, still to be tested at that time. It appears from the historical evidence that while Einstein was aware of Nordström's work on the scalar theory he had not appreciated that it was also a theory in a 'five-dimensional cylinder world'.

compactification radius Klein could reproduce the cylinder conditions of Kaluza. The emergence of electromagnetism in Kaluza's theory is also not a surprise anymore because compactification on S^1 induces essentially a $U(1)$ gauge-invariance in the theory (the symmetry group corresponding to S^1). This is a very important observation and allows the passage to the case where there is more than just one extra dimension. In fact, related to the $U(1)$ invariance of the five-dimensional theory, is the quantisation of charge, i.e. the Kaluza-Klein fields have electric charge quantised in terms of the mode number, n. It is tempting then to think of the $n = 1$ mode as the physical electron and identify the physical charge of the electron with the quantised charge of the Kaluza-Klein theory, thereby explaining the mysterious fact of charge quantisation. The problem is that this argument works only for modes higher than $n = 0$ and we have already seen that the higher modes have masses proportional to a very large scale. This internal contradiction in the five-dimensional theory was the reason for the loss of interest in it though it was later realised that it is possible to circumvent these problems in a theory with more than one extra dimension. The other reason, and probably more important from a historical point of view, for the loss of interest in Kaluza-Klein theory was the discovery of nuclear forces. The realisation that, other than electromagnetism and gravitation, there were two other fundamental interactions in nature was a major setback to the Kaluza-Klein programme. At that stage in the history of physics, it was more important to arrive at an understanding of these forces than worry about unification. It was a long and tortuous journey that the subject had to go through before the discovery of the Standard Model as the theory of strong, weak and electromagnetic interactions was made.

1.3 Non-abelian generalisations

The Standard Model is a gauge field theory based on the gauge group $SU(3) \times SU(2) \times U(1)$. Of course, this understanding came after a full 40 years of research in which time not only the phenomenological details of the Standard Model were understood but the importance of gauge theories as an important class of relativistic field theories was fully appreciated and its dynamics understood at the quantum level. This class of theories includes not only the abelian $U(1)$ theory of which electromagnetism is an example but also non-abelian gauge theories needed to understand the larger symmetries of strong and weak interactions. In retrospect, it is ironical that the first explicit non-abelian gauge theory (or Yang-Mills theory, as it is now called) was written down in the 1930s by Klein in an attempt to go beyond the five-dimensional Kaluza-Klein theory. Klein considered [18] the compactification of two extra dimensions on a sphere S^2. The isometry group of the sphere being $SU(2)$, the resulting theory was an $SU(2)$ gauge theory instead of electromagnetism. While this result did not impact the development of non-abelian gauge theories, which followed a completely different historical course, it was an important mathematical result and showed the way of generalising Kaluza-Klein theory to include larger gauge groups by considering compact manifolds of higher dimension with larger isometry groups. The particularly attractive feature of this is the group of isometries generated by the Killing vectors of the compactified space manifest as the non-abelian Yang-Mills symmetries in four-dimensional space.

Though there were earlier attempts, it was with the work of Cho and Freund [19] that the complete derivation of the four-dimensional gravitational and Yang-Mills theory from a higher dimensional theory was presented. The problem with the very attractive idea of obtaining Yang-Mills theories as isometries of the higher dimensional space is that it necessarily gave only curved space solutions for four dimensions and a flat-space solution was ruled out. Much of the effort that followed, pioneered by the work of Cremmer and Scherk, [20] was to try to achieve what was called a spontaneous compactification of the extra dimensions by including additional gauge and scalar fields. Solving the full Einstein equations in the presence of these additional fields yielded classical solutions in which the four-dimensional space was Minkowski.

1.4 Higher-dimensional supergravity

One of the most remarkable discoveries in high-energy physics in the 1970s was that of a new symmetry – supersymmetry. The generators of this symmetry are fermionic and obey a graded Lie algebra of anti-commutators. The product of two such generators is proportional to momentum, i.e. the generator of translations in spacetime, which makes it clear that supersymmetry is a new symmetry of spacetime. Because the generators of supersymmetry are fermionic, they relate fermions to bosons making thereby a fundamental connection between matter and force. Moreover, it is a symmetry that may well be realised in the world of high-energy particles, albeit not as an exact but as a broken symmetry for even when it is not exact it helps ameliorate the problem of severe fine-tuning that one encounters in the Standard Model in trying to compute quantum corrections to the mass of the Higgs particle. It is remarkable that since a product of supersymmetry transformations gives rise to a spacetime translation, by elevating supersymmetry to the status of a local symmetry one can generate local – i.e. spacetime-dependent – translations or general co-ordinate transformations. So gravity is a necessary component of local supersymmetric theories otherwise known as supergravity theories.

The simplest $N = 1$ version of supersymmetry involves only one set of generators but it is possible to make extended versions of supersymmetry. Noting that the supersymmetry generators change the helicity of a particle by half and that because of their anti-commuting property repeated application of the same generator on a given particle state will yield zero it is easy to see that with N sets of supersymmetry generators the maximal change in the helicity of a particle state can be $N/2$. If, following the informed prejudice in the subject, one then demands that there be no spins greater than 2 in the theory, then it implies that $N \leq 8$. It is possible to view these extended supersymmetries as $N = 1$ theories in higher dimensions and by counting degrees of freedom of particles in the super-multiplets it can be seen that the maximal $N = 8$ theory in four spacetime dimensions can be viewed as an $N = 1$ theory in 11 spacetime dimensions. This is what spurred interest in supergravity theories in higher dimensions.

If maximal supergravity theories exist in 11 dimensions, it was shown by Witten [21] that if one hoped to derive the Standard Model as isometries of a compact manifold, then the minimum dimensionality of the compact manifold is 7 – again a 11-dimensional spacetime! This led to the widespread belief in the 1980s that 11-dimensional supergravity was the ultimate theory because it held promise for the unification of all forces including gravity. Freud and Rubin then showed [22] that in 11-dimensional supergravity, the graviton (a second-rank tensor g_{MN} with 44 components in 11 dimensions) and the fermionic superpartner, the gravitino (a Majorana fermion with 128 components), could not form a complete supermultiplet. In order to match the bosonic and fermionic degrees of freedom it was necessary to introduce a massless antisymmetric tensor potential with three indices. Analogous to the case of electrodynamics, this potential gives rise to gauge-invariant field strengths with four indices. If the field strength or its dual is to have non-vanishing vacuum expectation value on the d-dimensional spacetime then Freund and Rubin showed that either d or $11 - d$ must equal the number of indices of the field, i.e. either $d = 4$ or $d = 7$. This argument has been spelt out in some detail here because it is a beautiful demonstration of how supersymmetry predicts the dimensionality of spacetime. Awada, Duff and Pope [23] studied this further by analysing the isometry group of S^7 viz. $SO(8)$, and relating it to $N = 8$ supersymmetry in four dimensions.

In spite of the early enthusiasm, 11-dimensional supergravity failed to deliver due to several problems which were encountered. One serious problem was the non-renormalisability of supergravity theories. Further, it was shown by Witten [21] that no compactification of an 11-dimensional theory using a compact seven-dimensional manifold can yield chiral fermions so crucial for constructing the Standard Model. None of the attempts to try to evade Witten's argument and generate a chiral fermion spectrum was satisfactory and the 11-dimensional model had to be eventually abandoned. Witten's observation, however, did not apply to compact manifolds of even dimensionality so attention now turned to supergravity theories in ten dimensions where chiral fermions could be generated. However, it is possible to view this theory as the zero-slope limit of a string theory and with the remarkable developments that were taking place in string theory, the drama moved to another stage.

1.5 Superstrings

String theory was originally proposed as a theory of strong interactions in the 1960s at a time when Quantum Chromodynamics was not yet formulated. At that juncture in the history of particle physics, a considerable amount of hadron-scattering data were available but a fundamental field-theoretic description was lacking. The apparent failure of field theory in describing strong interactions led to the suggestion that strong interactions should be understood in terms of S-matrix theory where the attempt would be to derive the dynamics of strong interactions from the properties of the S-matrix, such as analyticity, unitarity and crossing symmetry. Veneziano proposed a form of the scattering amplitude which satisfied these requirements and displayed the desired asymptotic behaviour and duality. The breakthrough in understanding the underlying dynamics of the dual model came with the realisation that this model could be obtained from a relativistic string in that the spectrum of hadronic states could be determined from that of a string. In spite of its initial success this string picture of the strong interactions ran into problems because of the persistent appearance of a massless, spin-2 hadron in the spectrum of states of a closed string which had no analogue in the hadronic spectrum and eventually the string model of hadronic physics had to be abandoned.

It was Scherk and Schwarz [24] who resurrected the idea of strings by suggesting a reinterpretation of string theory not as a theory of hadrons but as a theory of all interactions, which naturally incorporates gravity. The dreaded spin-2 massless excitation of closed-string theory was to be understood as a graviton, instead. The scale of string dynamics was also elevated to Planck scale from typical hadronic scales so that the corrections to classical gravity coming from string theory would be at appropriate short-distance scales.

We will be spelling out string theory in greater detail in Chapter 4, so here we will emphasise the main points of contact with Kaluza-Klein theories. The action for a classical relativistic string, which is either open or closed, can be derived from the requirement that it be proportional to the area of the two-dimensional surface swept out by the string. One can then proceed to quantise this system using standard techniques using some convenient gauge. It turns out that quantisation leads to an anomaly called the conformal anomaly making the theory inconsistent. There also appear negative-norm states in the string spectrum. Both these problems turn out to have the same solution: these disappear if the spacetime dimensionality is 26. A further condition also restricts the mass spectrum of the theory. But what is important is the fact that higher dimensionality of spacetime in string theory is not an *a priori* assumption but it emerges naturally from consistency requirements of the quantum theory of strings. Just as the Kaluza-Klein idea emerges naturally from string theory so does the idea of unification of fundamental interactions: the quantisation of the open and closed string proceeds along similar lines except that the open string has a spin-1 massless mode and the closed string has a spin-2 massless mode, suggesting a possible path to the unification of Yang-Mills theory and gravitation.

The quantisation described above is for the bosonic string. If we want to bring fermions into the picture it is done by supersymmetrising the action and bringing in Majorana fermions as superpartners to the bosonic co-ordinates. The quantisation proceeds as before but the consistency condition to get rid of the conformal anomaly yields $d = 10$, rather than $d = 26$.

Compactification of the extra dimensions to obtain consistent and realistic four-dimensional theories has been a problem that has received considerable attention. Compactification on manifolds of $SU(3)$ holonomy, i.e. Calabi-Yau manifolds, provides low-energy physics with $N = 1$ supersymmetry but chiral fermions continue to remain a problem in such models. Compactification on orbifolds may hold the key to getting chiral fermions. In spite of several promising attempts in these directions, it is fair to say that none of these are close to providing a realistic model of particle physics at low energies.

It was from a rather unexpected vantage point that the new onslaught on higher dimensions was launched about ten years ago. Much of this was due, at least in spirit, to new developments that took place in string theory in the mid-1990s. These new developments were mainly in the understanding of string theory at strong coupling and the progress was due to the discovery of a series of dualities in string theory which allowed strong-coupling and weak-coupling theories to be related to each other. These eventually led to proposals where the complete set of duals for all known string theories (with sufficient supersymmetry) was detailed. On the one hand, the dualities in string theory are like strong/weak-coupling dualities but, on the other hand, it is also intimately connected to electric–magnetic duality. In the latter case, while at weak (electric) coupling, electric charges appear as elementary quanta but magnetic monopoles appear as extended objects (solitons), in the strong coupling the roles get reversed and the basic quanta turn out to be magnetic. In string theory, similarly, duality gives rise to new solitonic solutions called branes. We will discuss this in more detail in Chapter 3 but the branes turn out to act like the familiar domain walls in field theory and serve as surfaces on which gauge and matter fields can be localised.

Arkani-Hamed, Dimopoulos and Dvali [25] considered a theory with $D - 4$ extra dimensions which has the Standard Model particles confined to a 3-brane (which is essentially a 3+1 dimensional surface) and only the gravitons are free to propagate in the full D dimensions. As usual, the extra $D-4$ dimensions have to be compactified to obtain the $3+1$ dimensional theory. But, since these extra dimensions are only 'seen' by gravity, these need not be compactified to length scales which are of the order of M_P^{-1} but it can be arranged that n of these extra dimensions are compactified to a common scale R which is relatively large, while the remaining dimensions are compactified to much smaller length scales which are of the order of the inverse Planck scale. Depending on the number of large extra dimensions, the magnitude of R could vary from a millimetre