International Journal of Theoretical Physics, Vol. 24, No. 4, 1985

The Development of Ideas in Twistor Theory

S. A. Huggett[1]

Received October 10, 1984

This paper presents a review of the main concepts of twistor theory. The emphasis is on the evolution of the subject from the original motivating ideas to the more recent work. In particular the physical and philosophical reasoning behind the use of the various mathematical structures is discussed.

1. PRELIMINARY CONCEPTS

Some of the most important motivations behind twistor theory can be traced back to the study of spin networks (Penrose, 1971, 1972a). A spin network is a combinatorial expression of the quantum mechanical procedures for combining nonrelativistic total angular momenta. In particular it represents a collection of particles exchanging angular momentum between themselves. Having said that, it is of course very difficult not to imagine these particles as being in some background space, but that would go completely against the philosophy of spin network theory, according to which the "background" space is *constructed* from the (purely combinatorial) spin network itself. Indeed it can be shown that it is possible to use parts of a network consistently to define the directions in Euclidean three-space corresponding to those particles with large angular momenta.

This result, although it is only partial (being a nonrelativistic scheme which also excludes displacements between the particles), fits very well into the attempt to question the validity of the use of two key mathematical concepts in theoretical physics. The first of these concepts—and the one whose basic role in physics is most obviously challenged by twistor theory—is that of the space-time point, and the second is the mathematical continuum. We are familiar with the elegance and power of the use of the space-time continuum in relativity theory (Penrose, 1968a) but its successes

[1]Mathematics Department, Plymouth Polytechnic, Plymouth, Devon, England.

0020-7748/85/0400-0391$04.50/0

should not prevent us from highlighting the physical irrelevance of some of its consequences (such as the fact that there are the same number of "points" in a volume of radius 10^{-13} cm as there are in the entire universe). This is especially important when we consider that the uncertainty principle together with mass energy equivalence prevents us from localizing a particle with arbitrary accuracy. Even if we were able to ignore quantum theory, however, it would still be unsatisfactory to use a mathematical model whose foundations are so unphysical. According to this philosophy, then, the objective is to extend the spin network result to a fully relativistic scheme. In the sense that space-time points become derived objects this is what twistor theory does, although as we shall see the approach is quite different.

Another area of research which was crucial in setting the scene for twistor theory was the study of zero rest mass fields and their conformal invariance (Penrose, 1965). It became clear from this work—and from the work on null hypersurface initial data done at more or less the same time (Penrose, 1980)—that not only was the two-component spinor formalism very powerful, it was also a more natural setting for these ideas than the usual tensor calculus. Consequently it was not long before spinors came to be regarded as more basic objects than tensors. This can be expressed mathematically in the form of the local isomorphism

$$SL(2, C) \to O(1,3) \tag{1}$$

Furthermore, the fact that spinors are null objects and the growing importance of the use of null hypersurfaces in general relativity led to the suspicion that the conformal group had an even more fundamental part to play in theoretical physics (Penrose, 1968b). Another significant aspect of (1) was that it indicated the value of the use of complex techniques. It was by no means the only instance of such a hint. It had for example been known for some time (Penrose, 1974) that the effect of a Lorentz transform on the celestial sphere of an observer is simply a conformal transform on the sphere thought of as a Riemann sphere. Secondly, the Kirchoff-type integral introduced in Penrose (1980) to evaluate a free field at any point in terms of the null datum for that field was reminiscent of a Cauchy integral. Thirdly, there was a strong feeling that the Fourier analysis method of describing positive and negative frequency could and should be replaced by a structure whereby functions on a real hypersurface in a complex manifold can under certain circumstances be considered as boundary values of functions defined to one side (positive frequency) or the other (negative) of the hypersurface.

Probably the most influential results (so far as complex techniques are concerned), however, came from the search for exact solutions of Einstein's equations. In the "impulsive wave" space-times (Robinson and Trautman, 1962; Penrose, 1972a) coordinates could be chosen in such a way that

Einstein's vacuum equations were reduced to the condition that a function on the wave front be *holomorphic* (whether the wave front was planar or spherical). It could also be shown that the null hypersurfaces containing the impulsive wave had to be shear free, so that the planar and spherical cases were the only two possible examples of pure impulsive gravitational waves. Furthermore both the real and imaginary parts of these holomorphic functions were important, which reinforced the conviction that complex geometry had a fundamental (if somewhat obscure) role in relativity. This takes us back to our discussion of the use of the mathematical continuum in theoretical physics, where the two formalisms of general relativity and quantum theory are based on the real and complex fields, respectively. It would clearly be an advance to use the same (complex) field for both, not least because complex analysis has (as we shall see) features which suggest that we may yet be able to move toward some form of combinatorial description.

At first sight it may seem that the obvious approach is to complexify Minkowski space. However, real Minkowski space is then nowhere near being a hypersurface in complexified Minkowski space (because it has four fewer dimensions instead of one), and it is therefore difficult to see how to incorporate the idea of representing positive and negative frequency in the terms mentioned earlier. Another more philosophical but no less significant objection to the complexification of Minkowski space is that such a technique would be applicable whatever the dimension of space-time. Given that we believe that space-time has four dimensions it is far better to seek a more specific mathematical structure which *only works* in four dimensions. Then a larger portion of the mathematical model is likely to be of direct physical relevance. This philosophy—which is a cornerstone of twistor theory—is also discernible behind the attempts to replace the space-time continuum altogether.

So if complexified Minkowski space is not the structure we are looking for what is? We must bear in mind that any new (presumably complex) formalism should at least take account of the following ideas: we would like to be able to think of space-time points as derived concepts, we expect conformally invariant objects to be important, and we would like to have a description of positive and negative frequency in terms of boundary values of complex functions.

The details of such a formalism were in part suggested by the work on null geodesic shear-free congruences by Robinson, who showed (Robinson, 1961) that given such a congruence a solution of Maxwell's free field equations could be obtained. As we shall see in the next section the crucial point here is that these congruences can be regarded as "half complexified" null lines.

2. EARLY TWISTOR GEOMETRY

Consider null geodesics in Minkowski space M. There are five (real) dimensions worth of them, because we can specify a null geodesic by (i) its intersection with some fixed spacelike hypersurface (three dimensions) and (ii) its direction along the null cone of that point of intersection (two dimensions). In other words the space of such null geodesics has five dimensions. The "fundamental theorem" of twistor geometry is that if we include the "half complexified" null geodesics (which we describe in detail in a moment) we obtain a *complex manifold* instead of merely a six-dimensional real manifold. This complex manifold is CP^3 (which is the space of four complex numbers up to an overall complex multiple) and we call it *projective twistor space* PT. The space of real null geodesics which we started with forms a five real dimensional hypersurface PN in PT.

The simplest description of these "half-complexified" null geodesics is in terms of complexified Minkowski space CM. There are two types of totally null complex 2-planes in CM, called α planes and β planes. An α plane is defined by solutions to the equation

$$z^{AA'} = z_0^{AA'} + \lambda^A \pi^{A'} \tag{2}$$

where $z_0^{AA'}$ and $\pi^{A'}$ are fixed and λ^A varies. The space of α planes in CM is exactly PT, and the correspondence between the two is summarized by

$$\omega^A = iz^{AA'}\pi_{A'} \tag{3}$$

To see this correspondence note that the pair $(\omega^A, \pi_{A'})$ (up to a complex multiple) defines a point in PT. If we fix that point then the solutions $z^{AA'}$ of (3) are given by (2). If $x^{AA'}$ is a *real* space-time point then the projective twistor

$$(ix^{AA'}\pi_{A'}, \pi_{A'}) \tag{4}$$

defines a *real* null geodesic (lying in the corresponding α plane) through $x^{AA'}$ in the direction $\bar{\pi}^A\pi^{A'}$.

One of the motivations behind twistor theory was that the concept of a space-time point should be derived, not primary. So the next task is to see how space-time points are represented in projective twistor space. We return to equation (3) but this time fix the (real or complex) space-time point $z^{AA'}$ and instead solve the equation for ω^A and $\pi_{A'}$. The solutions form a one complex dimensional projective space CP^1 (which we refer to as a line even though it has topology S^2) in PT. If the line lies in the hypersurface PN then the corresponding point is real. So points in PT represent α planes in CM and lines in PT represent points in CM. To complete the picture we must explain that planes in PT (in other words elements of PT*) represent β planes in CM, where a β plane is defined by solutions of (2) where $z_0^{AA'}$ and λ^A are fixed and $\pi^{A'}$ varies. We can see from (4) that a real null geodesic lies in exactly one α plane. It also lies in a unique β plane, so that a properly complexified null geodesic corresponds to an α plane and a β plane. This is why we referred to α planes as "half-complexified" null geodesics.

There are a couple more aspects of the geometry of twistor space we should discuss before moving on to some of the early results in the theory. Firstly, as can quite easily be seen from (3) two lines in PT intersect if and only if their corresponding points in CM are null separated. Secondly, the hypersurface PN in PT is defined in twistor terms by the Hermitian form (written in terms of nonprojective twistors)

$$Z^\alpha \bar{Z}_\alpha = \omega^A \bar{\pi}_A + \pi_{A'} \bar{\omega}^{A'} \tag{5}$$

where $Z^\alpha = (\omega^A, \pi_{A'})$. We have

$$Z = \{\rho Z^\alpha : \rho \neq 0 \in C\} \in \text{PN} \Leftrightarrow Z^\alpha \bar{Z}_\alpha = 0$$

For a more detailed discussion of basic twistor geometry we refer to Penrose and Ward (1980). We shall be content with the following summary of the relationship between space-time and projective twistor space:

PT	CM
point	α plane
line	point
plane	β plane
intersection of lines	null separation of points
line in PN	real point

Even at this early stage we are in a position to look back at some of the ideas leading up to twistor space to see what has happened to them. To begin with, the group of transformations of projective twistor space leaving the Hermitian form (5) invariant is $SU(2,2)$. This group is not compact and therefore has infinite-dimensional irreducible representations. So we do not expect any direct analogy with spin networks (where the underlying group was compact). On the other hand some of these infinite-dimensional representations (the zero rest mass free fields) will turn out to be very interesting. Also, we have the local isomorphism

$$SU(2,2) \to C(1,3) \tag{6}$$

which shows that, as expected, conformal invariance is fundamental to twistor theory. This is not surprising, of course, given that we constructed twistor space from null objects. Indeed it is only by singling out these null objects that we have been able to obtain the complex structure of twistor space. Furthermore, the usual way of thinking of a quantized space-time is to leave the points intact but quantize the metric. This leads to well-defined points and fuzzy null cones, which according to the twistor philosophy is exactly the wrong way around. If, instead, we quantize in twistor space we can expect to have well-defined null directions and fuzzy points. In particular spinors would survive, as indeed they should if the belief (expressed earlier) in their importance for the structure of space-time is to have any validity. This inversion of the usual roles played by space-time points and null directions is at the core of twistor theory and its power, which lies in the fact that it provides us with a new viewpoint, can be seen in all the subsequent work in the theory.

The first actual theorem expressed in twistor geometry was the *Kerr theorem*, which states that all shear-free congruences of null geodesics are given by the elements of sets of the form $\mathrm{PN} \cap Q$, where Q is some holomorphic surface in PT (Penrose, 1967). This theorem was at the same time encouraging and slightly worrying. It was encouraging because it demonstrated that a key aspect of curvature had been represented as a holomorphic condition in twistor space. It was worrying because when a shear-free congruence of null godesics goes through a region where the space-time is conformally curved ($C_{abcd} \neq 0$) the Sachs equations imply that it picks up some shear. Therefore conformal curvature seems to destroy the complex structure of twistor space. We will return to this problem later.

The next task which twistor geometry was asked to perform was the description of free zero rest mass fields. As we have seen, the study of these fields provided some of the original ideas for twistor geometry and yet at first glance their description in twistor space is a little daunting because a function on space-time is a function of *lines* in PT. A hint, however, was provided by the fact that a zero rest mass field is determined by its initial data, which is a function of three variables. In conjunction with the guiding philosophy that we should be looking for *holomorphic* structures this suggests that we consider holomorphic functions on regions in PT. (A function which is holomorphic all over PT has to be constant.) To get from such a holomorphic function to a zero rest mass field we first of all suppose that we wish to evaluate the field at the space-time point $x^{AA'}$. Then we restrict the function on PT to the line representing $x^{AA'}$, and recall that this line is actually a CP^1, which is topologically an S^2. If the region where the function is not holomorphic (called its singularity region) is such that it intersects the line $x^{AA'}$ in two disjoint patches then we can integrate the function along any contour separating these patches to obtain the value of the field at $x^{AA'}$ (see Figure 1). There will be lines for which these two patches have moved together, pinching the contour. These lines correspond to points in

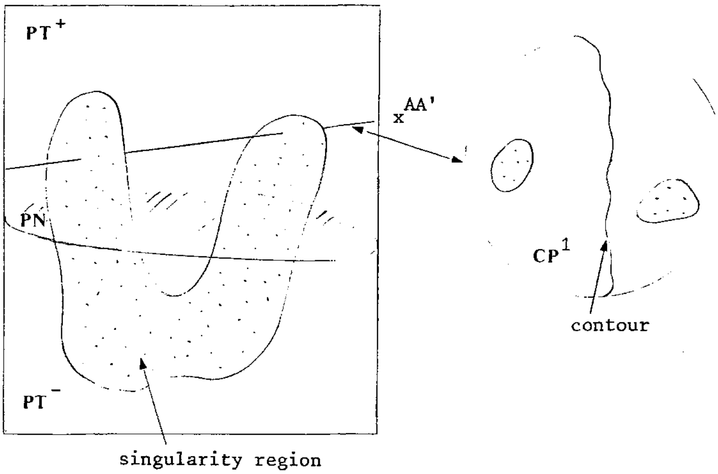

Fig. 1.

CM at which the field blows up. The importance of this procedure lies in the fact that the condition that the twistor function is holomorphic (in other words the Cauchy–Riemann equations) is necessary and sufficient for the corresponding field to satisfy the zero rest mass equations (Penrose, 1968b). While this result was undoubtedly a major step forward some difficulties remained. Firstly, it did not provide as neat a characterization of positive and negative frequency as had been hoped: a line lying completely in the upper half PT^+ of PT corresponds to a point lying in the future tube CM^+ of CM. If the field is finite for all such lines then it has positive frequency. In other words if the singular region of the twistor function intersects all lines lying in PT^+ in two disjoint patches then the corresponding field has positive frequency. Secondly, neither the twistor function nor the contour over which it is integrated are uniquely defined by the space-time field. This is because the integral is a Cauchy integral, and it has the unfortunate consequence that the collection of functions and contours corresponding to a single space-time field is rather ill defined.

Meanwhile twistor theory was still faced with the problem that the shear of a congruence of null geodesics is not in general preserved as one moves along the geodesics so that the complex structure of PT would not be expected to survive in a conformally curved space-time. One way out of this dilemma was suggested by the study of H spaces (Ko et al., 1979), which led to the concept of asymptotic twistor space, which is a twistor space defined relative to the null hypersurface $\mathscr{I}$ at infinity. This idea in turn developed into the theory of hypersurface twistors. Suppose S is a spacelike hypersurface in Minkowski space. Then we could have constructed PT by taking the space of intersections of α planes in CM with CS (the complexified hypersurface). In a general space-time $\mathcal{M}$ α surfaces will not exist, but we can still construct the space of "intersections" by considering complex curves in CS (satisfying certain conditions) to obtain a complex manifold P$\mathcal{T}(S)$ called projective hypersurface twistor space. The so far slightly obscure relationship between curvature and complex structure could then be clarified a little by the "hypersurface Kerr theorem." This states that a congruence of null geodesics in $\mathcal{M}$ is shear free at S if and only if it corresponds to some intersection P$\mathcal{N}(S) \cap Q$ where Q is a holomorphic surface in P$\mathcal{T}(S)$. So if S and S' are two hypersurfaces separated by a conformally curved region then P$\mathcal{T}(S)$ and P$\mathcal{T}(S')$ must have different complex structures. The detailed description of these structures, however, had to await further developments.

3. RECENT TWISTOR THEORY

Another important idea coming from the study of H spaces was that a complex space-time could be conformally right flat (Penrose, 1976), which means that the conformal curvature is

$$C_{abcd} = \psi_{ABCDA'B'C'D'}$$

The importance of such a space-time lies in the fact that α surfaces still exist, so that presumably the space of these α surfaces is a twistor space P$\mathcal{T}$ whose complex structure has been deformed away from flat twistor space PT. The next problem was to construct a space-time from a given deformed twistor space. In the flat case the points of space-time were the CP^1s in PT, and these CP^1s could be characterized by being compact holomorphic curves having the correct homology. It was by no means immediately clear that a four-parameter family of such curves existed in P$\mathcal{T}$, and the proof of this fact involved (among other things) a sheaf cohomological calculation. (We shall see in a moment that sheaf cohomology was soon to be introduced in another part of twistor theory.) This four-parameter family provided the points of a space-time and what was more it could also be proved that a space-time constructed in such a way was *automatically* conformally right flat, and that *every* conformally right flat space-time could be so obtained. In fact it was the conformal structure of the space-time which was obtained first. The metric, which followed soon after, needed a deformation of nonprojective twistor space considered as a bundle over the $\pi_{A'}$-spinor space, and was called the nonlinear graviton metric (Penrose, 1976). These startling results take us back to the original motivation for using the complex field in relativity, because we can now see in what sense complex analysis has "finite" or "rigid" features reminiscent of a combinatorial description. It was the requirement that the compact curves in $P\mathscr{T}$ be *holomorphic* which had the effect of singling out the finite-dimensional system of space-time points. If we had merely been looking at differentiable real curves we would never have been able to define anything other than an infinite-dimensional system.

In the last section we indicated the problems with the contour integral method of generating zero rest mass fields. In particular, a whole collection of twistor functions and contours was seen to correspond to one space-time field. This freedom in the choice of twistor functions and contours was found to correspond *exactly* to the freedom in choosing a representative cocycle for a sheaf cohomology class (Eastwood et al., 1981). In other words one space-time field corresponds to one sheaf cohomology class, so that we have (for example) the isomorphism

$$\begin{aligned}&\{\text{holomorphic solutions of } \nabla^{AA'}\psi_{A'\ldots L'}=0, \text{ where } \psi_{A'\ldots L'}\\ &\text{is symmetric in its } n \text{ indices and defined on } CM^{+}\}\\ &\cong H^{1}(PT^{+}; O(-n-2))\end{aligned} \tag{7}$$

Not only does (7) solve the problem of the accurate formulation of the collection of functions and contours corresponding to a given space-time field, it also provides the simple characterization of positive and negative frequency which had been sought. Indeed, the increasing use of cohomology theory has simplified and suggested new ideas in several areas of twistor theory.

One such area is the study of the "twisted photon" (Ward, 1977) in which anti-self-dual Maxwell fields can be described as a certain type of line bundle over PT. If vector bundles over PT are taken instead of line bundles then (so long as the vector bundles are trivial when restricted to CP^1s) solutions of the Yang–Mills equations are automatically generated. While this is another example of solutions of a differential equation in space-time being provided by holomorphic structures in twistor space it has in common with all the others the property that the differential equation is either self-dual or anti-self-dual. The major challenge for twistor theory is to discover how to solve Einstein's vacuum equations by first constructing a *left*-flat space-time in terms of PT and then somehow putting the two halves of the conformal curvature together. The work in progress on the first part of this program is in attempting to dualize the ordinary twistor description of a space-time point (Penrose, 1981). This is by no means the only current investigation in twistor theory, however. The theory as a whole is extremely rich and very effective at suggesting new directions for research (Hughston and Ward, 1979). We have been content here to concentrate on the development of the main lines of thought.

ACKNOWLEDGMENT

If I have been successful in writing an undistorted discussion of some of the ideas in twistor theory then my debt to Roger Penrose for introducing me to such a beautiful discipline will be clear throughout the paper. If not, then the responsibility is of course mine.

REFERENCES

Eastwood, M. G., Penrose, R., and Wells, R. O., Jr. (1981). *Communications of Mathematical Physics*, **78**, 305-351.

Hughston, L. P., and Ward, R. S., eds. (1979). *Advances in Twistor Theory*. Pitman, London.

Ko, M., Ludvigsen, M., Newman, E. T., and Tod, K. P. (1979). The theory of *H*-space, *Physics Reports*, **71**, 51.

Penrose, R. (1965). Zero rest-mass fields including gravitation: asymptotic behaviour, *Proceedings of the Royal Society of London*, **A284**, 159-203.

Penrose, R. (1967). Twistor algebra, *Journal of Mathematical Physics*, **8**, 345-366.

Penrose, R. (1968a). Structure of space-time, in *Battelle Rencontres*, C. M. DeWitt and J. A. Wheeler, eds., pp. 121-235. Benjamin, New York.

Penrose, R. (1968b). Twistor quantisation and curved space-time, *International Journal of Theoretical Physics*, **1**, 61-99.

Penrose, R. (1971). Angular momentum: An approach to combinatorial space-time, in *Quantum Theory and Beyond*, T. Bastin, ed., pp. 151-180. Cambridge University Press, Cambridge.

Penrose, R. (1972a). On the nature of quantum geometry, in *Magic without Magic: J. A. Wheeler*, J. R. Klander, ed., pp. 333-354. Freeman, San Francisco.

Penrose, R. (1972b). The geometry of impulsive gravitational waves, in *General relativity: papers in honour of J. L. Synge*, L. O'Raifeartaigh, ed., pp. 101-115. Clarendon Press, Oxford.

Penrose, R. (1974). Relativistic symmetry groups, in *Group theory in non-linear problems*, A. O. Barut, ed., pp. 1-58. D. Reidel, Dordrecht.

Penrose, R. (1976). Nonlinear gravitons and curved twistor theory, *General Relativity Gravitation*, **7**, 31-52.

Penrose, R. (1980). Null hypersurface initial data for classical fields of arbitrary spin and for general relativity, *General Relativity and Gravitation*, **12**, 225-264.

Penrose, R. (1981). Physical left-right symmetry and googlies, in *Twistor newsletter 12*, Oxford: preprint.

Penrose, R., and Ward, R. S. (1980). Twistors for flat and curved space-time, in *General Relativity and Gravitation, One Hundred Years after the Birth of Albert Einstein*, A. Held, ed. Plenum Press, New York.

Robinson, I. (1961). Null electromagnetic fields, *Journal of Mathematical Physics*, **2**, 290-291.

Robinson, I., and Trautman, A. (1962). Some spherical gravitational waves in general relativity, *Proceedings of the Royal Society of London*, **A265**, 463-473.

Ward, R. S. (1977). Curved twistor spaces, Oxford, D. Phil. thesis.