Proceedings of the Ninth International Conference on General Relativity and Gravitation

PROCEEDINGS OF THE NINTH INTERNATIONAL CONFERENCE ON GENERAL RELATIVITY AND GRAVITATION

JENA, 14–19 JULY 1980

Under the auspices of
International Society for General Relativity and Gravitation
International Union of Pure and Applied Physics
Friedrich-Schiller University, Jena

Edited by

E. SCHMUTZER

Department of Relativistic Physics, Friedrich-Schiller University, Jena, German Democratic Republic

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE

LONDON · NEW YORK · NEW ROCHELLE · MELBOURNE · SYDNEY

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia.

First published 1983

Printed in the German Democratic Republic by
VEB Druckhaus „Maxim Gorki", 7400 Altenburg

Library of Con[illegible]s catalogue card number:

ISB[illegible] 521 24669 5

Editor's Preface

The scientific material for the discussion groups of GR 9 was handed out to the participants of the conference in 3 Abstracts Volumes (about 750 pages) at the beginning of the congress.

This Proceedings Volume contains most of the GR 9 plenary lectures. Unfortunately the publication of this volume has been delayed by nearly half a year, because some of the authors could not keep the mutually agreed term for handing in their manuscripts, which was the end of September 1980. Although the publishers prolonged the time limit for submitting the manuscripts up to the end of 1980, some of the important contributions which were greatly appreciated by the participants have not arrived yet. I regret that these articles cannot be included in this volume. They are the lectures held by: I. Robinson, J. Ehlers, and Ya. B. Zeldovich. By reason of time the interesting lecture read by J. Trautman can only be published in form of an extended abstract.

In honour of Albert Einstein's 100th birthday in 1979 I — contrary to the tradition of the GR congresses — established a circle for historical studies on the occasion of GR 9 which was met with great interest by the participants. Both of the larger contributions to this topic read by H. Melcher and L. Pyenson are included in this Proceedings Volume.

For general information I present some data on the distribution of the participants per country and the grants given for GR 9. These facts are part of the Appendix, which ends with a poem on GR 9 by N. V. Mitskievich, who, unfortunately, could not take part in the congress.

The succes of GR 9 was guaranteed especially by the well founded contributions of the plenary lectures and the moderators as well as by the active support of Friedrich Schiller University Jena, the Ministry of Higher Education of the GDR, the IUPAP and the International Society on General Relativity and Gravitation (Berne). Their support is gratefully acknowledged. I express my sincere gratitude to the President of our Society, Prof. Dr. P. G. Bergmann, and its secretary, Dr. A. Held, as well as to the secretary of GR 9, Dr. R. Collier, and to our office secretary, Frau U. Kaschlik for permanent help in the organisation of the congress. I thank Miss C. Conlin and Herr E. Hahn for the translations. Last not least I would like to thank the staff of the VEB Deutscher Verlag der Wissenschaften, Berlin, for his sympathetic understanding and help in solving technical problems.

Jena, February 1981 — Ernst Schmutzer

Contents

OPENING SESSION

Welcoming Address
by the Rector of Friedrich Schiller University Jena

F. Bolck (Jena)

Ladies and Gentlemen,

I have the pleasure to welcome you in our country, in this town and especially at our University. — We are very happy about the decision of your Committee on General Relativity and Gravitation to give the 9th Conference of your Society to Jena, and that our University has the honour of being the host of this conference.

Your Conference has gained great interest which is expressed by the fact that the Deputy Minister for Higher Education, Mr Harry Groschupf, has come to this opening session. I welcome you, Mr Deputy Minister, and now I have the honour to inform you about an address to the Conference from our Minister of Higher Education, which I will read in German.

I would like to express my sincere thanks to you, Mr President, and to the members of your Committee for your kind interest in our University.

Our University was founded about 420 years ago in 1558. This was the time of the reformation. In its development our University has been closely connected with the names of many eminent scientists in the arts as well as in the natural sciences and medicine. This is not the place to go into historical details, but there are three important aspects which I would like to mention:

firstly: The role of the University in the time of German Classicism is connected with the names of Goethe and Schiller. Schiller was a professor of history in Jena and it is to him that the University owes its name,

secondly: The foundation of the Zeiss works is closely connected with the University.
Carl Zeiss was the University mechanician and Ernst Abbe was a lecturer at our University.
This combination was of great influence on the sciences, especially the physics at the University as well as on the developing optical industry.

thirdly: Our University was the first one in Germany to be reopened after the second World War in 1945. Since those days a new University has developed which is open to all people, who have as their aim to do scientific work in the interest of the population as well as in the interest of science itself and to educate the young generation in the spirit of peace.

It is well known that science can develop only in peace and that the application of scientific results postulates a permanent preservation of peace. Therefore the ideas of mutual understanding and peace are the aims of our country as well as of our University.

In connection with this scientific and social responsibility of the University theoretical physics plays an important role. Let me enter into a few remarks on some eminent representatives of this field:

It was Ernst Abbe who started with his thesis in 1863 a theoretical physical research of high international reputation in the field of optics. These investigations were the scientific basis of the industrial development in the Zeiss factory.

In 1912 F. Harres published his thesis about the Sagnac experiment where the light path nearly completely went through glass; and he approximately confirmed the theoretical hypothesis.

Georg Joos repeated the Michelson experiment in a basement of the Zeiss factory in 1929—30 with the confirmation of the special relativistic hypothesis; the limit of error was ± 1.5 km/sec. In those days this was the greatest precision of this experiment.

Between 1947 and 1951 Friedrich Hund and his coworkers carried out research work in relativistic quantum field theory.

Since 1957 Ernst Schmutzer and his group have started general relativistic research at our University. The research covers the following subjects:

- Unified field theory, especially projective relativity theory;
- Theory of the spinors in curved space;
- Quantization of physical fields in curved space-time;
- General relativistic continuum mechanics and thermodynamics;
- Electromagnetics and quantum theory in non-inertial frames;
- Exact solutions of the Einstein-Maxwell theory;
- the border area of gravitation and low temperature physics.

This research work has led to many publications and scientific conferences. But it also resulted in an interdisciplinary cooperation between physics and philosophy; these common efforts are concerned with the problem of the unity of physics. Ernst Schmutzer's publication "Relativity theory today — a contribution to the unity of physics" is also devoted to the philosophical interpretation of the relativity theory.

It is not possible to mention all the publications of Ernst Schmutzer and his coworkers, but about 10 books have been published on relativity theory.

I think I am correct, when I say that our history as well as the present scientific work have contributed to the decision to hold your 9th International Conference on General Relativity and Gravitation at our University.

I welcome you once more in Jena and I wish you successful scientific work and a pleasant stay in this town and its surroundings.

Welcoming Address
by the Minister of Higher Education of the GDR
(read by Rector F. Bolck)

H.-J. Böhme (Berlin)

Meine sehr verehrten Damen und Herren!

Der 9. Internationale Gravitationskongreß 1980 in Jena führt bedeutende Gelehrte und Spezialisten auf dem Gebiet der Relativitätstheorie und Gravitation aus zahlreichen Ländern der Erde zum wissenschaftlichen Gedankenaustausch und zur Diskussion neuester Forschungsergebnisse zusammen und läßt Impulse für die Aufhellung grundlegender physikalischer Zusammenhänge der Entwicklung des Kosmos erwarten.

Wir betrachten die Ausrichtung dieser bedeutsamen internationalen Tagung durch die Friedrich-Schiller-Universität Jena als einen wichtigen Beitrag der Deutschen Demokratischen Republik zur Entfaltung einer dem Frieden, den humanistischen Idealen und dem Menschheitsfortschritt verpflichteten Wissenschaft und als Würdigung der Leistungen unserer Wissenschaftler auf dem Gebiet der Relativistischen Physik.

Der 9. Internationale Gravitationskongreß in Jena ist Ausdruck der Wahrung, Pflege und Fortführung des progressiven wissenschaftlichen und humanistischen Erbes des Begründers der Relativitätstheorie, Albert Einstein, in der Deutschen Demokratischen Republik.

Die Begegnung von Wissenschaftlern aus allen Teilen der Welt wird einen Beitrag leisten zur Sicherung und Bewahrung des Friedens als Grundvoraussetzung für das weitere Gedeihen der Wissenschaften zum Wohle der Menschheit.

Ich wünsche dem 9. Internationalen Kongreß für Allgemeine Relativitätstheorie und Gravitation einen erfolgreichen Verlauf und den Teilnehmern schöne und erlebnisreiche Tage in unserem sozialistischen Staat.

Translation

Ladies and gentlemen!

The 9th International Congress on Gravitation unites eminent scientists and specialists in the field of General Relativity and gravitation from many countries of the world for the exchange of scientific ideas and the discussion of the latest results of their research, and new impulses for the elucidation of fundamental physical interrelations and of the development of the Universe are to be expected.

We regard the organization of this important conference by the Friedrich Schiller University of Jena as a considerable contribution of the German Democratic Republic to the development of a science obliged to peace, humanistic ideals and human progress and as an appreciation of the achievements of our scientists in the field of relativistic physics.

The 9th International Congress on Gravitation in Jena is an expression of the preservation, cultivation and continuation of the progressive scientific and humanistic heritage of the founder of Relativity Theory, Albert Einstein, in the German Democratic Republic.

The meeting of scientists from all parts of the world will contribute to the protection and preservation of peace as a fundamental prerequisite for the further development of science for human welfare.

I wish the 9th International Congress on General Relativity Theory and Gravitation successful work and the participants pleasant and eventful days in our socialist country.

Welcoming Address in the Name of IUPAP by the President of the National Physics Committee of the GDR

J. Auth (Berlin)

Mister chairman, ladies and gentlemen!

On behalf of the National Comittee on Physics of the German Democratic Republic I welcome all participants in the Ninth International Conference on General Relativity and Gravitation. I welcome you very heartly in our country. The physicists in the German Democratic Republic are very glad about the decision of the International Society for General Relativity and Gravitation to have their Ninth International Conference here in this country and in this town.

I welcome you also on behalf of the International Union of Pure and Applied Physics, the IUPAP, which is sponsoring your conference. Our National Committee is representing the IUPAP in this country. The German Democratic Republic is a member of the International Union of Pure and Applied Physics since 1960 and we think that the activities of this Union are very important for the advance in science, especially in physics, for a fruitful exchange of ideas, and for a useful cooperation in physics between the different member countries all over the world. And we think also, that the activities of this union can contribute to a lasting peace in the world. The conference we are opening now is a further contribution in this direction.

The German Democratic Republic is now thirty years old. We can say that with the foundation of the GDR a new stage began in the development of sciences in our country in general. The very successful evolution of sciences in the last thirty years is a convincing expression of the continuity of the science policy of the Socialist Unity Party of Germany and of the Government of the GDR that is derived from the laws of social development and based on the active link between working class and progressive science. Scientific basic research is governed in our country by 8 large long term national research programs in the main branches of science as mathematics, physics, chemistry and so on.

These programs have been worked out by the cooperation of very many scientists and the details of these programs, the main directions of research, the aims and the best ways to put them into practice are discussed in many consulting councils between scientists from the research institutes of the Academy of Sciences of the GDR, the universities and the industry. We see in this practice one special but important part of our real democracy. A very important characteristic of the physics research program is the very close connection of pure and applied physics. In this framework physics in our country is developing very well. You may see this very clearly at the physics department of the Friedrich Schiller University Jena, one of the leading universities in physics in our country and the friendly host of this conference.

And I think, that it is right to underline the fruitfulness of a close connection, a strong interaction of pure and applied physics, just at this place, at this town, as Jena in history of science and technology is one of the first places in the world, where this concept has proved to be fruitful in the very close relations between the Carl Zeiss Factories and the University in Jena.

But underlining the great social responsibility of our physicists and of physics in general for the solution of the great problems of mankind, as the energy problem for instance, does not mean, to see in physics only a practical instrument, a tool for solving problems. Physics is rather an extremely important, an essential part of our culture by itself. The great work of Albert Einstein has demonstrated this matter of fact once more in a very convincing way. Therefore it was only quite consequent, that the 100th anniversary of the birth of Albert Einstein one year ago in our country found a wide public interest. We are very glad to meet here in Jena again some of the famous scientists that took part in the governmental ceremony and in the scientific conference of the Academy of sciences and of the Physical Society of the GDR in 1979 in honour of Albert Einstein.

General Relativity and gravitational theory is one of the most fundamental parts of physics and any real progress in this field is of exceptional interest for physics as a whole. Unexpected connections may exist, and they may give us new proofs of the unity of physics and of the unity of the world. Gravitational waves, supergravity and other very important topics you will discuss in the next days here in Jena. And these discussions may give us some new understanding of fundamental problems of physics, some new insights into basic laws of nature that are controlling the evolution of the Universe.

The Ninth Conference on General Relativity and Gravitation in this sense is not only a very important point in our scientific life, it is of considerable philosophical interest and also an outstanding cultural event. It is a great honour and an obligation for us, too, the Physical Society and the National Committee on Physics of the GDR, that this conference takes place in Jena. I hope it will be a full success and you will have a nice stay in our country.

Opening of the Conference by the President of the International Society for General Relativity and Gravitation

P. G. Bergmann (Syracuse, N. Y.)

According to the official program, it is my task, as the retiring president of our Society, to open GR 9, the Ninth International Conference on General Relativity and Gravitation. To do so, I should need mostly a trumpet, plus the ability to play one. Lacking both I have thought about how to use best the few minutes consigned to this ceremonial act.

As relativists, and as physicists, astronomers, and mathematicians with at least a tangential concern for general relativity and its ramifications, we all are conscious of the Albert Einstein Centennial Year, which has just come to a close and which has been the occasion for many of us to join hands in celebrations all over the world, commemorating the contributions that Albert Einstein has made to science, and to the social betterment of humanity as well. The conference GR 9 will in many respects serve as a continuation of the gatherings last year. We shall share recent scientific advances in general relativity and related areas, and we shall do so as an increasingly international group, including, for the first time, a colleague from the People's Republic of China.

The year 1980 marks another anniversary, the twenty-fifth year after the first international conference on general relativity, which had been conceived by Wolfgang Pauli to mark the fiftieth anniversary of (special) relativity and which had been welcomed and supported by Einstein in the months preceding his death. This conference, at Berne, had been attended by fewer than ninety colleagues. GR 9, by contrast, has drawn somewhere between five hundred and one thousand participants, the precise number not yet known to me. The list of members of the Berne conference of 1955 is to be found in the famous Fourth Supplement of Helvetica Physica Acta. Looking it over we find the names of many who have continued to be active in our field, and, sadly, also the names of those who are no longer with us, foremost among them the president of the Berne conference, Wolfgang Pauli, and, our most recent loss, Christian Møller, who had been the first president of the International Society for General Relativity and Gravitation.

Other colleagues this morning will present you with their views as to the present status of our field, but I shall take this opportunity to mention at least what seem to me among the most exciting aspects of relativity research just now. The principal contribution of general relativity to the edifice of physics appears to me the recognition that space and time form part of the physical universe, and that their geometric properties must be explored by the methods of experimental science. The state of our comprehension may well be cast into the form of axioms, and one possible axiom of this kind is that space-time be a pseudo-Riemannian manifold, or perhaps, instead, that it be a Weyl-type conformal manifold, or something else. But any such axiom is a proposed law of nature, a hypothesis concerning the actual properties of space-time, and its claim to validity must be viewed in that context.

Astronomy has presented us in recent years with new relativistic "laboratories", and the newest are pulsars in doublestar systems. I have no doubt that so far astronomers and space scientists have barely scratched the surface of the new possibilities, and that soon we shall be confronted by avalanches of new discoveries.

On the theoretical level the most exciting development to me is that a number of early attempts at a unitary field theory appear to converge with approaches stimulated by elementary particle physics. For a number of decades it appeared that general relativity was an appropriate physical theory for astronomical systems and in cosmology, but that it had nothing to contribute to the physics of the very small. It now looks as if this will change, and we find that ideas originated at one end of the scale stimulate progress at the other end. Perhaps what are fond hopes today will mature into viable theoretical structures within our lifetime.

It is often said that science is an international enterprise, which requires for its health the free circulation of discoveries and ideas throughout the world. We all consider this statement a truism, which requires no elaborate justification. But science is a human enterprise, interacting in many ways with social and political developments. If you agree with me that human survival depends on the maintenance and the continuing strengthening of peace among the nations, then obviously we scientists bear a special responsibility for doing our part toward this end.

In forming our International Society we have attempted a new model that deviates from the standard pattern exemplified by the Scientific Unions: We encourage individual membership, and we accept corporate membership only by scientific and academic organizations, rather than by country. It remains to be seen whether this kind of international organization serves well its technical purpose, the support of research in relativity in our case, and also fulfills its further task of strengthening the bonds of international friendship. I certainly hope so.

The holding of international meetings is beset with all kinds of problems and difficulties, among which the raising of the necessary funds is but one. Those of us who have been concerned with the preparatory work are aware of the tremendous work and effort that has been exerted first of all by Ernst Schmutzer and by R. Collier, both of the Friedrich Schiller University of Jena, our host institution. I have no doubts that their efforts have been aided by officials of the University, by its Rektor, Professor F. Bolck, and by the younger members of the local relativity group. They in turn have been supported by various authorities of the Government of the German Democratic Republic, the DDR, and we owe thanks to all of them. Finally I should mention the very great help by the Secretary of the Society, Dr. Alan Held, who managed to act as the international problem-solver of GR9 in many ways while seeing to its fruition the major contribution of our Society toward the Einstein

Centenary, "General Relativity and Gravitation", in two volumes, which has come out just a few weeks ago, and of which a few copies will be found here for all to look at.

On this happy note of accomplishment I shall close. GR 9 is now officially open!

Address: Prof. Dr. Dr. h. c. Peter G. Bergmann
Department of Physics, Syracuse University
Syracuse, New York 13210, USA

OPENING LECTURES

Einstein's Second Century

J. A. Wheeler (Austin)

Predict our progress in the century now beginning? That, we know, is beyond our power. We ask ourselves instead what *tasks* and *opportunities* lie ahead.

The Illusion Behind Reality and the Reality Behind Illusion

To see how far we have yet to go in the exploration it may help to recall how far we have come. Two hundred years ago electricity was a toy. A hundred years ago we were barely beginning to use electricity. Of the electron we did not even know the name, much less the existence. Today we see a great railway train, a thousand tons of mass, thunder over a bridge, driven by electrons, colorless abstract objects, running smoothly, quietly along overhead wires. Those electrons in turn derive their energy from the world of the fantastically small, from billions upon billions of quiet acts of nuclear fission, fission yielding neutrons that we cannot see or touch or hear; and those neutrons in turn yielding more fission, plus the heat that makes power.

As if we did not have in electrons and neutrons a world of enough abstraction and untouchability, we get our daily news by a carrier that is no particle at all. It flies in to us on the wings of an invisible wave, the flip-flop bending of electric lines of force, lines that are born and die without ever once in their existence acquiring anything that anyone would call separability, identifiability, touchability.

Have we not discovered in the past century that physics is a magic window? Is it not the destiny of physics to show us the illusion that lies behind reality — and the reality that lies behind illusion? And is our task not therefore immensely greater than we once thought it to be?

A great textbook of physics of a hundred years ago gives the impression that it would be enough for making our science complete to know the list of all substances and the physical properties of each. Today it is popular to speak of our task in equally positive terms: to know the three or four or five laws of force and the principle linking them. Will this vision in time to come prove itself equally short sighted?

Can any vision of science that confines itself to science ever reveal the scope of science? The modern university has a finance officer who knows how much money to give to physics, how much to mathematics, and how much to philosophy. However, at the time this world of ours came into being, this finance officer was not around to keep his kind of order. The act of creation became a great tangle of philosophy, mathematics and physics.

To straighten out today and tomorrow our understanding of the deeper issues, can we count on our friends, the philosophers? "No", they tell us, "We don't know enough physics to do the job. It is more reasonable for you physicists to learn the necessary philosophy." Can we leave the task to our wonderfully helpful colleagues, the mathematicians? With our immediate work they help us more with each passing decade; but not ordinarily with the finding of the central questions, the formation of concepts, and the fitting together of the pieces of the larger puzzle. On those enterprises is the impression justified that mathematics and philosophy even retreat before the advance of physics? How else could they better tell us what they see our future task to be?

In Einstein's first century physics won some understanding of physics. In Einstein's second century physics has a far greater task. It must win some understanding of existence itself.

Einstein told us that one question concerned him more than any other: "Did God have any choice in the creation of the world?" How else better than by that question can we symbolize the greatness of the tasks and opportunities that lie before us?

Jena, Leibniz, and the Issues Raised by Leibniz

No one who wanders about Jena, reflecting on the challenges of our own day, can fail to be reminded of the great men of Jena's past. They walked and talked together on these very streets: the Friedrich Schiller from whom this university gets its name; his friend, also poet and thinker, Johann Wolfgang Goethe; and, above all, and in a still earlier day, Gottfried Wilhelm Leibniz.

Leibniz, one of Einstein's great heroes, more clearly than anyone was intellectual great-grandfather of Einstein's lifetime question, is there room for "any choice in the creation of the universe" — and of other queries tributary to that central issue: absolute space? absolute time? and space and time as preconditions for knowledge?

Space, Time and Inertia

Leibniz questioned the Newtonian concepts of absolute space and absolute time. Ernst Mach translated this doubt into concrete form. Acceleration relative to absolute space? What else could that mean but acceleration relative to the frame of reference defined by the faraway stars? No conception did more than this "principle of Mach" to drive Einstein to his 1915 and still standard "general relativity" or "geometrodynamics". How else could Einstein have conceived of accounting for gravitation, not as a foreign and physical force propagated through space, but as a manifestation of the curvature of space?

Does Einstein's theory really in the end automatically contain within it a proper mathematical formulation of Mach's principle? Does it explain how "mass-energy there fixes inertia here"? Or is this an antiquarian idea that has outlived its usefulness? So many for a time supposed. Then came the first investigations on the so-called initial value problem of general relativity by Élie Cartan, André Lichnerowicz and Yvonne Foures-Bruhat, followed by the still more useful results of the latter (now Yvonne Choquet-Bruhat), James W. York, Jr., Naill O'Murchadha and others. Thanks to them, new light broke on Einstein's equation.

We came to see general relativity as part of dynamics in the great tradition. It deals with the dynamics of geometry. We have only to tell the equation what the geometry of space is now — and how fast that geometry is changing — and it will tell us what the geometry of spacetime itself is everywhere and at all times. But what that spacetime geometry is at a point, Einstein tells us, is law and measure for what inertia is at that point.

To "determine inertia here for all time" we thus have to "specify the geometry of space everywhere at one time". But this specification, now, is only then possible when we state where all the sources of mass-energy are, now. This is the sense in which we learn at the end that inertia here is fixed by mass-energy there.

Space and Time as Prerequisites for Knowledge

Having in hand a modern version of Mach's principle, we have come to appreciate better the motives of Leibniz in questioning the doctrine of absolute space. But Leibniz has even more in mind when he warns that "... space and time are orders of things and not things". So does Einstein when he adds, "... time and space are modes by which we think and not conditions in which we live".

Neither statement is idle talk. Both are calls to action. They demand of us a foundation for space and time. That underpinning even general relativity does not provide. To this demand Immanuel Kant makes an heroic response. He points out that no knowledge whatsoever is possible unless we have in hand the tools to combine in thought the detached elements of experience. He is not concerned with how these sensory impressions reach us. However, to organize them — or any experience — in any meaningful way, he stresses, would be absolutely impossible in the absence of the two essential conditions for sense perception: space and time. Time and space exist because knowledge exists!

As science advances, Kant's considerations lose favor. Relativity contributes to this fall. Out of "*a priori* pure reason" does Kant purport to derive the existence of space as one category and time as another? And does Einstein demonstrate that there are no such things as space, separately, and time, separately, but only spacetime? Then there must be something wrong with Kant's reasoning.

It is one position to say that Kant is all wrong. It is quite another to say that he is right in general conception, wrong in detail. If he could publish evidence in 1754 arguing for a slow change in the axial rotation of the earth, and in 1755 calculate how fast the Milky Way must rotate to withstand gravity, perhaps he had as good a nose for physics as anyone today. Few conceptions are grander than Kant's idea that space and time are preconditions for a knowable world. Properly to assess that view — and, more broadly, to answer the questions of Leibniz — is business that is unfinished — and important.

What "Pregeometry" Lies Behind Geometry?

A less ambitious question often received attention a few decades ago, "Why does space have dimension three?". Nowadays we are beginning to raise our sights and ask a question that is bigger though not so big as Kant's. We ask, "How does the world manage to give the impression that it has dimension three?"

Nowhere more clearly than at a crack does a crystal reveal that it is not a continuum at the submicroscopic level. Nowhere more conspicuously than at a selvedge does cloth show that it is not a continuum but woven out of thread. Spacetime — with or without "gauge" or "phase" or "internal spin" degrees of freedom — often considered to be the ultimate continuum of physics, evidences nowhere more clearly than at big bang and at gravitational collapse that it cannot be a continuum. Obliterated at the bounds of time, we see no escape from concluding, is not only matter, but the space and time that envelop that matter.

If a crystal is built out of electrons and nuclei and nothing more, if cloth is woven out of thread and nothing more, we are led to ask out of what "pregeometry" the geometry of space and spacetime are built.

Nothing is more basic to the description of a crystal than elasticity, and nothing is less basic. There is no such thing as "elasticity" in the space between the electron and the nucleus. A hundred years of the study of elasticity would never have revealed that it goes back for its origin to electrons, nuclei and Schrödinger's equation and nothing more. The direction of understanding went, not from the large to the small, but from the small to the large. If elasticity is the last place to look for a clue to Schrödinger's equation, geometry would seem to be the last place to look for a clue to pregeometry — and quantum theory the first. It is no wonder that "quantum geometrodynamics" or "quantum gravity" is such a central topic of concern to so many today.

Space and time: reality or illusion? Whatever Einstein's second century has in store for us on this question, Leibniz will surely still have the last word: "Although the whole of this life were said to be nothing but a dream and the physical world nothing but a phantasm, I should call this dream or phantasm real enough, if, using reason well, we were never deceived by it."

Gravity as Illusion

The illusion behind reality, and the reality behind illusion, show up nowhere more strikingly than in gravity itself. In fig. 1 the ball is thrown across the room. The action of gravity makes its path parabolic, or so we are accustomed to say. Fig. 2 illustrates that same ball thrown across the same room with the same initial velocity.

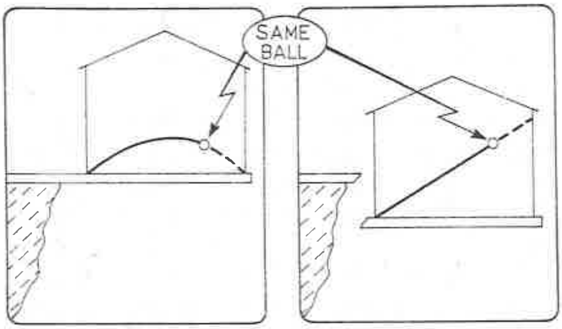


Fig. 1 Fig. 2

Now its path is straight. The change is not in the ball but in the frame of reference from which it is regarded. The room has been dynamited loose from the cliff. No longer is it driven from its natural track through space by the thrust of the beam. Instead it is in a condition of free fall or, better stated, "free float". In a freely floating frame of reference gravity disappears. Einstein abolished gravity as the first step toward explaining gravity. A fifteen-year old granddaughter expressed the idea in these words,

"What's the fault of the force on my feet?
What pushes my feet down on the street?
Says Newton, the fault is at the earth's core.
Einstein says, the fault's with the floor.
Remove that and gravity's beat" Frances Ruml (1978)

or as Stephen Schutz characterized Einstein's geometric theory of gravitation in response to a 1966 examination question, "Rather than have one global frame with gravitational forces we have many local frames without gravitational forces." Acceleration, the old measure of gravity, is recognized as illusion. That acceleration depended on choice of reference frame. Relative acceleration of nearby test masses, the new measure of gravity, is reality.

Reality, Illusion and Quanta

What we shall call illusion, and what reality, has proved even more difficult to clarify in quantum physics than in gravitation.

We walk through the art gallery on our way to visit again a favorite picture. We pass by the painting "Impressions", first shown by Claude Monet in 1863 at the Salon des Refusés. To a glance so brief how does the canvas project a coloration so vivid? Because quanta are so numerous. So a simple calculation informs us. A single dab of paint in the single second of our passage throws into a single pupil of our eye 50,000 photons. These quanta of light, this stream of information, is so impressive that there's no mistaking it. No wonder that our predecessors could believe light to be continuous.

In actuality, we know today, light consists of individual quanta of electromagnetic energy. Moreover, each of these photons is accidental in its time of emission and direction of travel. This Einstein taught us in the same magic year, 1905, in which he gave us special relativity. To him more than anyone we owe the lesson that "God plays dice".

Unteach this lesson, circumvent indeterminism, overthrow complementarity, restore predictability: that was the directly opposite thrust of Einstein's endeavors from the late 1920's to the end of his life. When Richard Feynman, then a graduate student, was developing in 1941 his beautifully simple distillation of quantum mechanics into a "sum over histories", I called on Einstein to describe the progress of the work. Every conceivable classical history contributes to the probability amplitude for a transition with the same weight as every other history. That is Feynman's principle of the "democracy of histories", I explained. However, these contributions, identical in magnitude, differ in phase one from another. The resulting destructive interference wipes out the contributions of all but histories close to the classical history. How could nature operate more beautifully, I exclaimed. Does this not make you more willing to accept quantum theory? "No", Einstein replied, "I still cannot believe that the good Lord plays dice" — and then added in his humorous way, "Of course I may be wrong; but perhaps I have earned the right to make my mistakes".

To the extent that chance comes in, Einstein knew, predictability goes out. What else is reality if it is not predictability? We can understand why he told Otto Stern, "I have thought a hundred times as much about the quantum problems as I have about general relativity theory". We can also appreciate why Einstein, in his famous 1935 paper with Boris Podolsky and Nathan Rosen, made his central objection to quantum theory this, that in his view it conflicts with any "reasonable definition of reality".

Your concept of reality is too narrow — that was the thrust of Bohr's reply. It — and elucidation of the term "phenomenon" — marked the climax of the three-decade-long dialog between the two men. In the early years of that debate no issue brought more insight than the double-slit experiment (fig. 3).

The illumination level can be made so low that the record made by each individual photon is subject to registration. Does that photon arrive *via* both holes in the doubly slit metal sheet — or only one?

The accumulation of many photons builds up on the photographic plate the traditional pattern of interference fringes. The result would seem to argue that each photon comes through both holes. This conclusion discomfited Einstein. He pointed out that in principle one can measure the lateral kick each photon gives to the photographic plate — to the right if it arrives from the left hand hole; to the left, if from the right. The photon arrives from both holes? And yet arrives from only one hole? What inconsistency! Is not quantum theory logically self-contradictory? To establish this point was Einstein's endeavor in the first phase of the dialog.

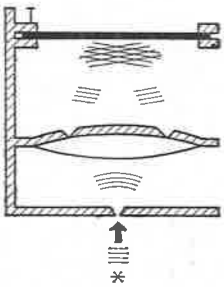


Fig. 3

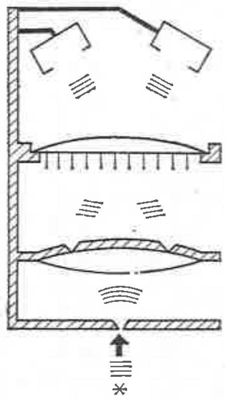


Fig. 4

Bohr's reply is well known. To register the interference fringes we hold the photographic plate fixed; for example, by inserting the pin shown in fig. 3, upper left. We have to remove the pin and let the plate slide freely to record the lateral kick imparted to the plate by the photon. Put the pin in or take it out: we can do either, but not both at once! We can observe fringes, evidencing that photons have made use of both holes. Or we can observe the kick, telling us from which hole the photon arrived. But no experiment can display the two features of nature at the same time. Complementarity, Bohr tells us, is a governing principle of knowledge: "Any given application of classical concepts precludes the simultaneous use of other classical concepts which in another connection are equally necessary for the elucidation of the phenomena."

After Bohr had successfully defended the logical consistency of quantum theory, Einstein in the final phase of the great debate attacked the theory as incompatible with any reasonable idea of reality. We do not have to turn to the EPR experiment of 1935 to meet the central lesson, Bohr's concept of "phenomenon". We see it as simply as anywhere in today's so-called "delayed-choice" version of the double-slit experiment (fig. 4).

The photographic plate has been sliced into strips like the blades of a Venetian blind. They stand open in the diagram. Above is a new lens. It focuses onto one photocounter a photon that comes through one hole in the doubly slit metal sheet; onto the other counter, a photon that comes through the other hole. Alternatively we can flatten the blades to make the photographic plate operative. Then we can register one more contribution to the usual pattern of interference fringes.

In the delayed-choice experiment we use a timed source. The photon it emits we permit to pass through the doubly-slit metal sheet. Only then do we decide whether to open or to close the blades of the Venetian blind. Open? Then we learn through which hole the photon came. Closed? Then we augment our interferometric record of photons that came through both holes. But the photon had already gone through the metal plate at the time we made our choice. Are we not therefore in this delayed-choice experiment deciding what shall have happened *after* it has *already* happened?

"Phenomenon"

"Decide what shall have happened after it has already happened" is the wrong set of words. "Phenomenon" is the right word. Bohr had to introduce it to make clear his position *vis à vis* Einstein on the EPR experiment. We can use that word to state the central lesson of quantum physics: "No elementary quantum phenomenon is a phenomenon until it is a registered phenomenon."

Specifically, it is wrong to speak of the "route" of the photon in the double-slit experiment. It is wrong to attribute a tangibility to the photon in all its travel from the entrance slit to its last millimeter of flight. A phenomenon is not yet a phenomenon until it has been brought to a close by an irreversible act of amplification such as the blackening of a grain of silver bromide emulsion or the triggering of a photodetector.

Nature at the quantum level is not a machine that goes its inexorable way. What answer we get depends on what question we put, what experiment we arrange, what registration arrangement we choose. The disposition of the equipment is inescapably involved in bringing about that which appears to be happening.

Quantum Phenomenon: Elementary Act of Creation?

How did the universe come into being? Is that some strange far-off process, beyond hope of analysis? Or is the operative mechanism still going on today?

Of all the signs that testify to "quantum phenomenon" as being the elementary act of creation, none is more striking than its untouchability. In the delayed-choice double-slit experiment we could have intervened at some point along the way with a different measuring device, before the phenomenon to-be had become a phenomenon. Nevertheless, that intervention would have given us no more right than before to say what the photon was doing in all its long course from point of entry to point of detection. Yes, we would have had a new phenomenon. No, we would have come no closer than before to penetrating to the untouchable interior of the phenomenon. For a process of creation that can and does operate anywhere, that reveals itself and yet hides itself, what could one have dreamed up out of pure imagination more magic — and more appropriate — than "quantum phenomenon?"

Of all the features of that "act of creation" that is the elementary quantum phenomenon, the most startling is that seen in the delayed-choice experiment. It reaches back into the past in apparent violation of the normal order of time. The distance of travel in the laboratory double-slit experiment might be three meters and the time ten nanoseconds; but the distance can as well be billions of light years and the time billions of years. Thus, the decision in the here and now to observe "from which hole" or alternatively to register "interference of the contributions from both holes" has an irretrievable consequence for what one has the right to say about a photon that was given out long before there was any life in the universe:

> „Sein Licht braucht eine Ewigkeit
> Bis es dein Aug' erreicht!
> Vielleicht vor tausend Jahren schon
> Zu Asche stob der Stern,
> Und doch steht dort sein milder Schein
> Noch immer still und fern."
>
> Gottfried Keller

Gravitational Lens and Delayed Choice

Two astronomical objects, known as 0957 + 561 A, B (fig. 5), once regarded as two distinct quasistellar objects or "quasars" because they are separated by six seconds of arc, are considered now by many observers to be two distinct images of one quasar.

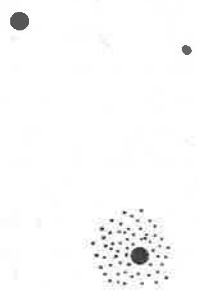


Fig. 5

Evidence has been found for an intervening cluster of galaxies, and in particular for one galaxy G-1 naturally identified as the principal component of a gravitational lens, roughly a quarter of the way from us to the quasar. Calculations by several groups indicate that a normal galaxy at such a distance has the power to take two light rays, spread apart by some sixty thousand light years on their way out from the quasar, and bring them back together at the earth. Thus there is nothing in principle to prevent the promoting of the double-slit delayed-choice experiment from the scale of the laboratory to the scale of the cosmos.

The natural stages in this evolution would seem to be (1) gross determinations of distances and speeds, (2) interferometric measurements via radio emissions from the quasar source, and (3) interference at the single quantum level. Each in turn is worthy of a little consideration.

If and when one of the quasar images indicates a flareup, then the other image should also briefly brighten, but a month or so later — according to calculations of C. C. Dyer and R. C. Roeder — and for a simple reason. The two routes of travel, extending over billions of light years, are bent by slightly different amounts, and therefore have slightly different lengths. Whithin a few years, surely, our colleagues in the world of astrophysics will detect such flareups, measure the flare-flare interval, and eventually even determine how that interval alters from year to year in consequence of quasar drift and lens motion.

Today thin strands of measurement connect us with the most remote objects we know in the universe. Tomorrow will not those strands thicken to a network? And as astronomy, patient spider of the stars, continues the spinning of this web, will it not grow to be even more spectacular and revelatory than the marvellous grid which geodesists have laid down on the earth to track the shape and drift of continents?

As photography reveals the flare-flare interval, and its rate of change with time, it will set the stage for radio interferometry. That interferometry will compare the phase of the radio waves recorded on two tapes. Those tapes will not contain, as today, records taken at two points of one source ("very long base-line astronomy"), but records taken at one point of the two apparently different sources produced by gravitational lensing. The delay and Doppler shift of one as compared to the other will allow one to determine still more precisely the galactic grid and its drift.

Radio interferometry operates at the classical level. So many photons come into play that quantum effects don't show up. However, one can hope to work someday with a part of the spectrum where only one photon at a time comes into consideration. On the way one will face an unprecedented problem in delay-line technology. One will have to learn how to keep alive, by superconductor or otherwise, and for a chosen time, of hours, days or months, the "image-A probability amplitude". One will also have to be able to bring it back out of storage with the right time delay, and right Doppler shift, to give it a determinate phase relation to the "image-B probability amplitude". Then the choice is clear (fig. 6): Operate the receptor, with its two photocounters, so as to tell "by which route" the photon comes. Or insert the half-silvered mirror, $(^1/_2)S$, and, with sufficient counts, determine with arbitrary precision the phase difference between the two routes.

The first such cosmic-scale delayed-choice experiment lies in the future. Today quantum theory already foretells its lesson: We have no right to say that the photon travelled from the quasar of red shift $z = 1.41$ to our delayed-choice detector by route A or route B or a linear superposition of the two. We are dealing with an elementary act of "creation". It reaches into the present from billions of years in the past. It is wrong to think of that past as "already existing" in all detail. The "past" is theory. The past has no existence except as it is recorded in the present. By deciding what questions our quantum registering equipment shall put in the present we exercise an undeniable choice in what we have the right to say about the past.

The phenomena called into being — by choices of what measurements to carry out — reach backward in time in their consequences as symbolized in fig. 7, back even to the earliest days of the universe. Registering equipment operating in the here and now has an undeniable part in bringing about that which appears to have happened. Useful as it is under everyday circumstances to say that the world exists "out there" independent of us that view can no longer be upheld. There is a strange sense in which this is a participatory universe.

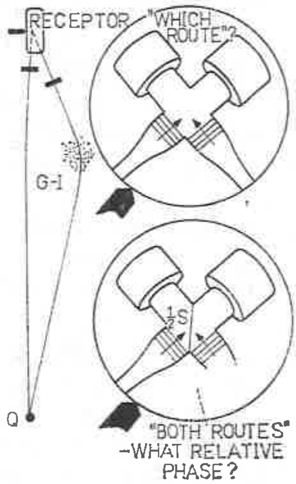


Fig. 6

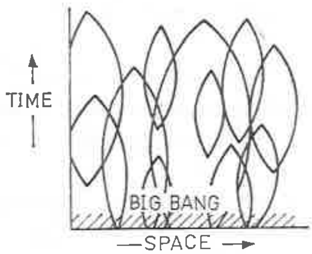


Fig. 7

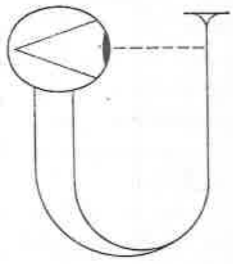


Fig. 8

The universe, depicted schematically by the letter "U" in fig. 8, starts with a big bang (upper right), grows (thickening calligraphy of the U), and ultimately gives rise to one and another registering device, symbolized by the eye. That recorder signals by an irreversible act of amplification the arrival of a photon of the primordial cosmic fireball radiation. Only then has the "long travelled" photon finally led to an elementary quantum phenomenon, a happening that one person can communicate to another in plain language. In such acts as this a tangible reality is at last conferred upon the past. The past so inferred is remote and hot. In that heat, registering equipment could not survive, let alone come into being. Neither could there be any such thing as meaningful communication.

It takes the present to give us a meaningful past, even as it took the past to give us the present. The past, we come to recognize, is an elaborate plaster-work construction of imagination and theory trowelled onto a sparser framework of observation, observations symbolized by the lattice work of fig. 7.

As theory — in the shape of general relativity — plasters deeper and deeper back into the past, it finds itself fabricating a troublesome architectural feature, an outburst of creation, a big bang, a "singularity." Is that singularity, we have to ask, anything but a plasterer's device to accomodate — and, having accomodated, conceal — the billions upon billions of elementary quantum acts that frame existence? What is the one big act of creation but illusive screen over many little acts of creation? What is the theoretical untouchability of the big bang but sum and deceptive substitute for the undeniable untouchabilities of all those individual elementary quantum phenomena? Where else than in them does that transformation come about — from the nothingness of the undecided to the somethingness of the decided — that makes the world?

Any more challenging question than this to light up Einstein's second century it would be difficult to name. We can recapitulate it in a triple phrase: "Must build, do build, how build." The universe declares, "Must build! I do not endure forever. There has to be a way for me to come into being." Quantum phenomenon replies, "Do build! Granted a piece of equipment, an experimental device, a probe, that asks me a question at the elementary quantum level, I provide out of nothingness an answer. Without supplementary parameters, without hidden variables, without any known internal machinery, out of indefiniteness I build definiteness — at the microscopic level." That declaration and that reply leave for us this question, "How build — at the microscopic level? What architectural principle offers itself for combining these many little definitenesses — the reality — into the world's all-encompassing definiteness — the illusion?"

Einstein's first century explodes Newton's absolute space and absolute time — with gravitational forces — into many local spacetime frames without gravitational forces. Einstein's second century has to explode "global reality" into many microscopic realities.

The many local Lorentz frames Einstein united into one curved spacetime manifold. The arbitrariness of the coordinates at first seemed to take all definite structure away from everything we call solidity. In the end physics, after being moved bodily over onto the new underpinnings, shows itself as clear and useful as ever. Now we are required to move physics a second time, over onto the still more ethereal foundation of elementary phenomena. We cannot doubt that the second transplantation will give us a still more successful physics.

The first move took away gravitation as a foreign and physical force transmitted through space. It gave us back gravitation as a manifestation of the curvature of space. The second move will surely take away much more — space and time themselves, as well as fields and particles — and give them back to us in a new light and new language, transfigured. To transform our outlook will be a great creative enterprise. Surely contributory to that undertaking is much of the work going on today, in many places, by many colleagues. About some of this research we shall be hearing in this conference. It includes investigations on the production of particles out of the vacuum, on vacuum polarization, and on quantum fluctuations in the geometry and topology of space. Quantum gravity continues to give us new insights into the dynamics of space geometry, also the meaning and limitations of the concepts of "spacetime" and "time". Time, we are more and more coming to realize, is not a primordial category in the description of nature. It is secondary, approximate and derived. Surely no analysis will ever be able to unravel the structure of existence which does not transcend the category of time.

Turning to the topic of time at a humbler level, we celebrate the many alternative ways of treating time, and the evolution of geometry with time, that have been developed in recent years. Among them we may pick out one by way of illustration, and an application of it that throws some light on our larger problem. Variously called "extrinsic time", "York time", "mean curvature", "trace of the extrinsic curvature", "trace K", "Tr K", and simply, "contraction", it measures the fractional decrease of space volume per unit of time.

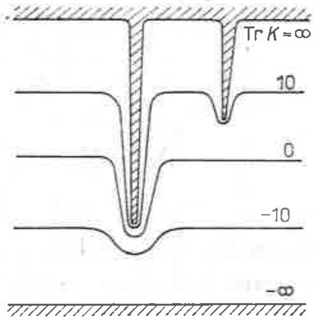


Fig. 9

A wide class of closed model universes, according to J. B. Marsden and F. Tipler, admit a unique "slicing" or "foliation (fig. 9) into a sequence of space-like hypersurfaces." On each hypersurface the "contraction time" is constant. The value of this constant differs from slice to slice. The contraction approaches $-\infty$ near one end of time, the big bang, and $+\infty$ at the other, gravitational collapse. This time parameter is useful in analyzing an interesting question about collapse.

In a closed model universe what is the architectural relation between (1) the big crunch of the entire geometry and (2) the black hole formed by collapse of some of the matter? For a simple model we may take a "dust-filled" Friedmann 3-sphere universe, cut out of it — with a "2-sphere cookie cutter" — a fraction, say 1/120 of its volume and replace it with a cloud of dust of slightly higher density, itself a small sector of a Friedmann geometry. Starting in this way near the big bang, the larger and the smaller sectors both expand. However, the smaller and denser sector goes through its cycle of expansion and contraction in a shorter time. It collapses to a black hole. Fig. 10 shows it on its way to this fate. Surrounding it is an empty space. There the geometry has the Schwarzschild character characteristic of any spherically symmetric center of attraction. The slicing of the global 3-geometry by spacelike hypersurfaces of constant contraction time shows these hypersurfaces engloving more and more closely a single spike, like one of the two spikes in fig. 9. That spike, that black hole singularity, is not touched by a single one of these hypersurfaces except the final one, the one of infinite contraction, of final collapse. In this way A. Qadir and I have been led to conclude that the black-hole singularity and the big crunch are not two separate features of the geometry, but part and parcel of a single, global, singularity — renewed incentive for concern about what happens to physics at the gates of time.

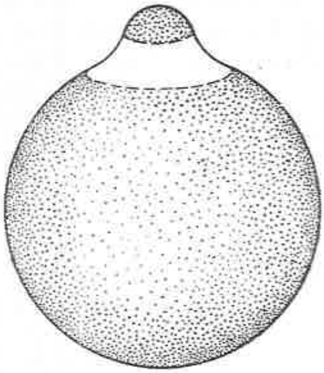


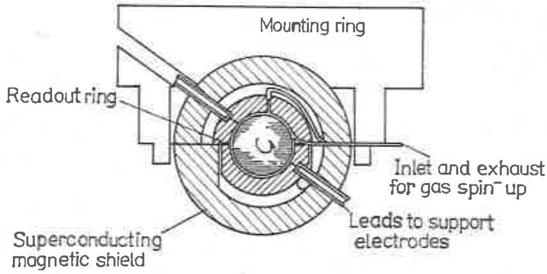
Fig. 10

In our subject there are challenges and opportunities for us all, from those concerned with concepts to those who love mathematics, and from those excited by fields and particles to those attracted to astrophysics and experiment. The rich colleagueship of those with all these interests makes gravitation physics what it is today:

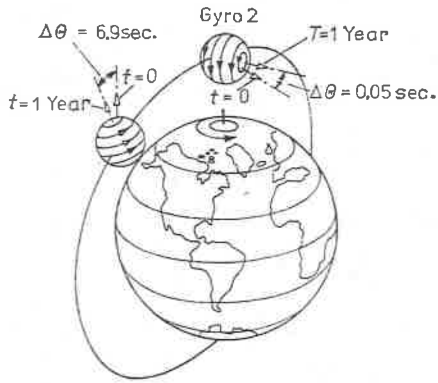
„Zum Werke, das wir ernst bereiten,
Geziemt sich wohl ein ernstes Wort;
Wenn gute Reden sie begleiten,
Dann fließt die Arbeit munter fort."

Friedrich Schiller

One cannot mention the word "experiment" without recalling the interest we all feel in the ongoing search of C. W. F. Everitt and W. Fairbank and their colleagues (fig. 11) to detect and measure the "gyrogravitational force" generated by the rotation of the earth (fig. 12), an effect as different from everyday gravitation as magnetism is different from electricity. Neither can we put out of our minds the many other tests of relativity receiving attention today, and among them especially gravitational radiation. The pulse produced by the collapse of a star in a distant galaxy is predicted to produce a fantastically small alteration in lengths here on earth. Weak though the wave is in this sense, it nevertheless contains so many quanta that it admits without any question a description in terms of the concepts of classical general relativity. It is different with the response of the measuring device. It is so weak that there is doubt that it can be measured, or measured as precisely as one would like, without pushing measurement technology — and measurement theory — into the quantum domain. There, issues are encountered to which V. B. Braginsky has given the name "quantum non-demolition" and with which he, A. B. Manukin, C. Caves, K. S. Thorne and others have grappled, with profit to our understanding. The pioneering going on, both theoretical and experimental, will surely have payoffs for measurement in fields far removed from gravitation.



Gyroscope for relativity experiment Fig. 11



Relativistic precession of gyroscope in 500-mile polar orbit Fig. 12

In no way does physics make a more powerful contribution to the larger community than in the instruments it provides for everything from biology to medicine and from astronomy to manufacturing. Not otherwise can one understand how it comes about that, quite apart from the variety of instrument makers and the variety of sizes in which instruments come, mankind now makes daily use of over two thousand kinds of instruments.

Scientists are delegated by society to keep watch for all mankind on what lies ahead of peril and promise. Whether we work on better instruments or better mathematics or better theory, we all count ourselves as friends of the future. As friends, we are prepared to go anywhere, see anyone, ask any question that will help us make progress with our work. No words are greater inspiration in our enterprise than Einstein's, "In my opinion, there is *the* correct path and... it is in our power to find it."

Perparation for publication assisted by University of Texas Center for Theoretical Physics and by NSF grant PHY 7826592.

Address: Prof. Dr. John A. Wheeler
Department of Physics, RLM 5.208
University of Texas at Austin
Austin, TX 78712, USA

Prospects for Relativistic Physics

E. Schmutzer (Jena)*)

1. Retrospect

In opening the 9th International Conference on General Relativity and Gravitation here in Jena today, we should thank such people as W. Pauli, A. Mercier, A. Lichnérowicz, M.-A. Tonnelat — to name but a few — who after the terrible turmoils of the Second World War and following the "Golden Jubilee Conference" in Berne 1955, found the energy to start the series of the international GR congresses [1]. I myself have been fortunate enough to be able to take part in these international meetings of scientists, with one exception, since Warsaw in 1962. Not only did I gain from these congresses professionally, they were also milestones of international understanding. I am convinced that Jena 1980 will also serve both components — science and humanism. It is in this sense that I would like to give our guests from 51 countries a warm welcome. I hope that you find the congress fruitful and wish you many memorable personal meetings, especially during our cultural events. If we give this basic idea priority during our congress we will be doing justice to the great heritage of the past master and founder of our science, Albert Einstein, whose second century we are inaugurating here in Jena.

Furthermore, I would like to offer all plenary lectureres, moderators and the many hardworking helpers from Jena, particularly the Secretary for GR9, Dr. R. Collier, and our office secretary, Mrs. Kaschlik, my most heartfelt thanks for all the work they have done so far and are doing this week. I should further especially like to thank the GDR Ministry for Higher Education, the IUPAP, the Gravity Research Foundation and the Secretariat of our Society for the generous financial assistance and the administration of the Friedrich Schiller University for their continuous support throughout the conference preparations.

The content of our congress is expressed in our conference logo: The R symbolizes Relativity as the basis of the new theory of the form and structure of the fundamental physical laws of nature acting in 4-dimensional space-time.

In my philosophical interpretation the G should instead of "general" rather symbolize the phenomenon of gravitation which has been recognized as the geometrical curvature of space and time. Gravitation is the focus of our present-day specific research. Our conference symbol is so designed as to acknowledge Newtonian gravitation — characterized by the symbol Φ — as the first step of cognition on the way to Einsteinian gravitation. In the sense of the continuity of scientific development

*) in memoriam Dr. med. Christel Schmutzer

without the Newtonian gravitational theory the more advanced Einsteinian gravitational theory would not exist.

If we draw up a comprehensive balance of general-relativistic physics today, we can state with satisfaction that in view of the framework of the limits of validity of its theoretical structure it can be regarded as having been outstandingly confirmed.

a) General-relativistic Einsteinian mechanics has proved itself excellently. It is — as the perihelion and periastron motions show — in a position to explain very small deviations in the motions of celestial bodies from the prediction of Newtonian mechanics.
b) General-relativistic Einsteinian gravitational theory can be regarded as experimentally satisfactorily confirmed in spite of the minuteness of effects on its main fields of application with an uncertainty of 1% and less.
c) General-relativistic cosmology has afforded an outstanding insight into design and structure of our Universe. Within its framework the phenomena of world expansion and of microwave background radiation are explained as a matter of course.
d) General-relativistic Maxwell theory covers the phenomena of electromagnetism in curved space-time in a logically satisfactory way.
e) General-relativistic Dirac theory signifies a decisive step towards describing quantum phenomena in curved space-time.

We consider that the general-relativistic Maxwell theory and Dirac theory still contain considerable potential which has not been fully exploited experimentally. An even better utilization of the interaction of gravitation with electromagnetic and quantum phenomena ought to prove very useful in various areas of experimental Relativistic Physics, especially in proving the existence of gravitational waves.

Let me now ask: Where do the unsolved problems of the present day theoretical Relativistic Physics lie? If we disregard many unsolved questions of detail, relating to the proven foundation, and if we keep the main line of development in view, we are still lacking, as you all know very well, the logical fusion between physics in curved space-time and quantum physics. Or in other words: the logical fusion of Einstein space-time and Hilbert space. Our own conceptions of this process were presented in several papers five years ago which dealt as a first stage with quantum mechanics in arbitrary frames of reference. We also sketched quantum field theory in rough outlines [2].

In the light of this challenge to fuse relativity theory and quantum theory it is legitimate to ask whether the series of GR conferences will stop at some time in the future and a qualitatively new series of RQ conferences will start.

After this general retrospect we would now like to go into more detail.

2. The experimental verification of the Einstein theory

Experimental Relativistic Physics has been receiving great impetus for about ten years. About one third of the activities of the GR conferences are today devoted to this field. Our conference will pay a great deal of attention to this tendency.

Terrestrial and planetary experiments

From the lectures by R. D. Reasenberg and I. I. Shapiro we will learn about the latest state of research on terrestrial and planetary experiments using radioastronomy. In summary one can say that Shapiro's radar echo method has provided a decisive contribution towards progress in the verification of Einstein's theory. It has become an indispensable device for measuring the radar time delay of electromagnetic waves in the planetary system (Shapiro effect), the deflection of electromagnetic waves by the sun and recently for measuring the perihelion motion by very precise ranging of planets and sun satellites. The effects just mentioned confirm Einstein's theory according to the state of research in 1979 with an uncertainty of 1% and better.

With the help of the Viking project, where extremely accurate maser-clocks are being used on an orbiter round Mars, the Einstein theory was confirmed last year up to an accuracy of 0.2%.

The projected goals for the Viking project of 1985 using highly developed radioastronomy methods consist of the following objectives:

a) to measure second order gravitational effects, where in the sense of an "experimentum crucis" competitive gravitational theories which offer partial explanations for first order effects should be put to the test;
b) to determine the mass quadrupole moment of the sun and the planets in order to gain new insights into the internal structure of these celestial bodies;
c) to introduce a serious search for planets of other fixed stars;
d) to investigate the detailed structure of galaxies and especially of their cores.

Furthermore, let me mention that the Mössbauer effect, which by using gamma rays, allows an extremely precise measuring of lineshift, is probably still the best experimental means for detecting frequency shifts of the electromagnetic waves in the gravitational field. It has verified the Einstein theory with an uncertainty of 1%.

Finally, let me note that in recent years the binary pulsar system PSR 1913 + 16 has become a very interesting object of research regarding periastron motion and gravitational radiation. The immense periastron motion of

$$(\Delta\varphi)_{\mathrm{PSR}\,1913+16} = 4.226^\circ$$

per year will still be of use for many conclusions in Relativistic Physics.

Exotic stellar objects and cosmology

The present state of research in exotic stellar objects and in cosmology will be outlined in the lectures by H.-J. Seifert and Ya. B. Zeldovich.

Even if we still do not possess an accepted and completely consistent quasar model, a considerable progress has nevertheless been achieved in the last few years. It is fairly certain that thermonuclear energy sources do not account for the enormous energy bursts of quasars. Furthermore the observed polarization of the electromagnetic waves is in contradiction to the radiation of thermonuclear processes. The synchrotron type of radiation observed, namely the polarization and distribution of the radiation over the spectrum from the ultraviolet to the radio region, points to the existence of strong magnetic fields, where the bursts of the radiating plasma chiefly follow the direction of the magnetic axes. I am quite sure that at our congress the hypothesis will be exhaustively discussed whether in the centre of a quasar there really does exist a supermassive black hole which could be the cause of the enormous energetic effects.

Superdense stellar objects have become a bonanza for Relativistic Physics. It is only through the relativistic gravitational theory in connection with relativistic electromagnetics that we have been in a position to supply the correct scientific basis for the physics of these objects with the observed mass densities. Now we possess a rich collection of empirical material on neutron stars (pulsars) in the optical and X-ray region. The theoretical expectations of relativistic magnetohydrodynamics for the magnetic fields of pulsars of about 10^{12} Gauss were outstandingly confirmed in 1976 by the measurements of the research group led by J. Trümper and others. It would be the task for a complex research group consisting of nuclear physicists, elementary particle physicists and relativists to trace the equations of state of matter at such extreme densities, pressures and temperatures.

The most interesting objects in this field of cosmic physics are black holes. Perhaps we will hear this week something new about the situation in the experimental discovery of black holes whose thermodynamics was essentially established by St. Hawking. The question whether black holes exist or do not exist has become crucial to the Einstein theory of gravitation with respect to the limits of validity of this theory. If black holes really exist with their topology in nature, then the Einstein theory is valid far beyond the region to which many researchers would like to restrict it. In this case — in connection with the wellknown questions of singnature — we would come across entirely new discoveries about the internal dimensionality of 4-dimensional space-time. If black holes with these properties do not exist in nature, then we have come to the limits of validity of the Einstein theory and must start looking out for a higher theory. In this connection — though not being inclined to agree with his basic line of thought — I recall the warning words of Ch. Møller about the "breakdown of physics" which was said to be induced by Einstein's theory, and of his endeavours over a period of more than 20 years to establish a tetrad gravitational theory.

It turned out that more Abstracts than average have been submitted to the discussion group led by N. Rosen about alternative gravitational theories. So we can say that many research activities today are situated at what could be a theoretical nodal point.

In the field of cosmology we can take as granted that the rough structure of the Universe from the moment of time a few seconds after the big bang can be very well described by the Friedman model of the expanding universe. We should always be aware of the fact that findings such as world expansion and microwave background radiation which confirm the Friedman model belong as organically connected facts to the best evidence of the Einstein theory.

During the next few days we will hear how far the hitherto conventional numerical values for the Hubble factor and the temperature of the background radiation:

$$cH \approx 60 \text{ km/sMpc}, \qquad T \approx 2.65 \text{ K},$$

can be further refined.

After sifting extensive statistical material M. Koch [3] gave the value

$$q = 0.64$$

for the acceleration parameter as very probable, which would correspond to a closed universe.

Gravitational waves

As has been usual for the last ten years, our congress will also devote a great deal of time to the problem of gravitational waves. This problem will be treated in various plenary lectures. J. Ehlers will deal with this subject from a theoretical angle in connection with the mechanical equations of motion. J. E. Marsden will probably refer to this subject in dealing with the gravitational initial value problem.

The lectures by V. B. Braginsky and K. S. Thorne will give us a general survey of the present state of experiments in detecting gravitational waves, and L. P. Grishchuk will deal with the perspectives of electromagnetic detectors. In the five corresponding discussion groups we will have the opportunity of learning the details of gravitational wave experiments. It is to be expected that the three main experimental tendencies: Weber cylinders, Braginsky monocrystals and electromagnetic detectors will complement each other very well.

In recent years another very important research area has been added to these three approaches: research by means of quantum detectors, an area whose theory has among other things to take into account the influence of gravitation on superconductivity.

Let me finally mention that additionally to the direct proof of gravitational waves by detectors there was developed in recent years the indirect astrophysical method which was exemplified by the binary pulsar PSR 1913 + 16 mentioned above with extreme accuracy. The new value

$$\dot{P} = (1.04 \pm 0.13) \cdot \text{quadrupole result}$$

recently found by J. H. Taylor for the temporal alteration of the revolution period of the pulsar certainly brings us a considerable step forward in evaluating the Einstein-Eddington quadrupole formula. Time will tell whether this system really radiates according to the quadrupole formula

$$Q_{\text{Qu}} = \frac{\varkappa}{8\pi \cdot 45c} \dddot{D}_{ab} \dddot{D}^{ab}$$

for the gravitational radiation power
($\varkappa$ Einstein's gravitational constant, D_{ab} mass quadrupole moment).

3. Some unsolved fundamental theoretical problems

A survey of theoretical relativistic research reveals many unsolved problems. This opens wide perspectives to our subject and it is at the same time an appeal and a challenge particularly to the brightest of the younger generation. In our remarks we can only touch on the tasks which seem to us to be most important and we have already referred to what might be the limits of the Einstein theory with respect to superdense mass configurations.

Exact solutions

Certainly we all agree that in view of physical inconclusiveness of the linearised Einstein theory it is correct to emphasize the field of the exact solutions of the Einstein-Maxwell equations. Many scientists have been researching in this field. In Jena we have been working in this direction for more than a decade. The collected experience of research available has been published in the monograph "Exact solutions of Einstein's field equation" by D. Kramer, H. Stephani, E. Herlt (Jena) and M. A. H. MacCallum (London). The plenary lecture of D. Kramer and H. Stephani will underline a few main points here.

For some time the investigation of exact solutions of the Einstein-Maxwell theory has been enriched by the step towards the complex solution methods which I. Robinson will report about and by involving highly developed computer methods about which R. A. d'Inverno will speak to us.

For which physical problems is the study of exact solutions of fundamental importance? Let us pick out a few of these, not in order of importance:

Certainly a complex of questions regarding gravitational radiation plays a central role here. We ought to arrive at exact and conclusive statements about the radiation field of bound mass distributions with internal degress of freedom of motion. Of course we can only expect mathematical solutions for the most simple configurations.

As an example the exact solution of the gravitational two-body problem would be of fundamental importance. We are aware that the solution of this problem is very difficult. However, international research should be continued, because the solution of this problem would also have enormous significance for cosmogony. It seems to us that up till now we are not able to state conclusively whether — presupposing gravitational radiation — the twobody problem sun/earth or earth/moon (as an idealisation) would end in escape or collapse.

As a first step in the direction sketched it would even be of immeasurable value to physics if we had an exact solution for the interior Kerr problem for a rotating fluid body — in analogy to the interior Schwarzschild solution.

Quantization in curved space-time

If we disregard the mathematical and technical difficulties, we can state that we have a mathematical grasp and in principle a physical understanding of the quantization of the non-metrical fields in Minkowski space using the rectilinear Galilei coordinates. The quanta of these fields are the elementary particles. Further, the situation becomes more complicated if we consider the same elementary particles, objectively existing in Minkowski space, in curved coordinates or even from a non-inertial frame of reference. As these last two tasks do not involve any interference with objective physical facts, the difficulties occurring must be regarded as being of mathematical nature.

We are confronted with a qualitatively new situation if these quanta are situated in a gravitational field. And this is actually the normal case: Our devices are set up in the gravitational field of the earth. They measure the properties of elementary particles which exist as quanta in the gravitational field. This means, the quantization of non-metric fields in a gravitational background field is, therefore, a genuine physical challenge which has been on principle understood and has also been mathematically solved for simple cases. In his lecture G. W. Gibbons will report on the state of research on this subject.

If we continue along this line of thought we arrive at the basic question: Is the gravitational field also quantized in nature? In other words: Do gravitons exist in nature as quanta of the gravitational field? The understanding of this problem has been hotly contested up to now.

If one allows oneself to be too strongly guided by the analogy with non-metric fields, then the quantization of the gravitational field appears to be a quite natural consequence. The mathematical theory of this was worked out in detail by P. A. M. Dirac and many others.

But we can also have serious doubts about this analogy, as we know that the gravitational field represents in quite decisive points basically quite a different kind of physics. We are going to hear more about this from D. Brill and P. S. Jang. What, for example, is the significance of quantizing the metric in a rotating carousel in Minkowski space, where there is no curvature and therefore no gravitation? Here we come up against basic questions of physics to which we can still give no clear answer. It is doubtful whether bimetric theories which attempt to separate the "coordinate-metric" from the "gravitation-metric" will bring genuine progress here. It seems to me that at this point we strike the hard fact of the principle of equivalence between kinematics and gravitation. It is only when one begins to doubt this basis of Relativity that one obtains room to manoeuvre for the quantization of gravitation.

If on the other hand, we adhere to the classicity of gravitation which is underlined by many arguments, then we are apparently still confronted by the old contradiction in the Einstein equation:

$$\underbrace{R_{ik} - \frac{1}{2} g_{ik} R}_{\text{classical}} = \underbrace{\varkappa T_{ik}}_{\text{quantic}} .$$

It has sometimes been suggested that in order to avoid this contradiction the vacuum expectation value should be placed on the right hand side:

$$T_{ik} \to \langle 0 \, | T_{ik} | \, 0 \rangle .$$

But this is an ad-hoc interference with the essence of the theory considered and is hardly compatible with self-contained Lagrange-Hamilton formalism.

Supergravity

The failure of Einstein's idea of a unified field theory on the basis of the geometrization of gravitation, electromagnetism and possibly of other fields as well, has resulted in a stagnation in this branch of physics lasting for many decades. However, I believe that most theoreticians find Einstein's philosophical basic idea attractive — but admittedly in contrast to W. Paulis's thesis: "What God has put asunder let no man join together".

Going back to an old idea of H. Weyl, the gauge field theories have brought a new theoretical stimulus. It is a well-known fact that for a field system consisting of a matter field and a Maxwell field a suitable combination of gauge transformation (for the Maxwell field) and phase transformation (for the matter field) guarantees the gauge-phase-invariance of the whole field system. If we now look at a rather more special case of a free matter field system which is only invariant with respect to a global phase transformation, then the transition from the global to the local phase transformation leads us directly to the Maxwell field as a compensating field. As is well known, this consideration is also the basis for introducing the Yang-Mills fields as compensating fields.

It is still a controversial question whether this analogy equally applies to the gravitational field which one would like to consider as a compensating field. (Meanwhile the designation "gauge fields" has been adopted for the compensating fields of the type I described, although hardly anything has remained of the original gauge idea.) Here the gravitational field is accounted for as the compensating field which is required to compensate the effect which occurs when one makes the transition from global tetrads to local tetrads $h_{(i)k}(x^j)$ which are attached to the metric as follows:

$$h_{(i)k}h^{(i)}{}_l = g_{kl}.$$

With this idea we are really entitled to doubt whether this procedure really does include gravitation and not just the coordinate effects.

Going beyond these gauge considerations it appears that in recent years some fundamental theoretical progress has been made by the discovery of the supersymmetry of boson and fermion fields. The point here is that field systems which are constructed from boson and fermion fields exhibit invariance properties by the combined boson-fermion transformation (in abstract notation):

$$\tilde{B} = c_1B + c_2F,$$
$$\tilde{F} = \gamma_1B + \gamma_2F$$

which then in the sense of the Noether theory supply the basis for the understanding of conservation laws.

This supersymmetry is the basis for the theory of supergravity as a unified field theory. In this way it would be possible to give a unified treatment of the gravitational field and of the electromagnetic field as boson fields and of the spinor fields of the elementary particles as fermion fields, so that the way would be opened for the fusion of the Einsteinian gravitational theory with the elementary particle theory — if this new research trend should prove true. The plenary lectures by P. van Nieuwenhuizen and S. Ferrara will acquaint us with the most up-to-date state of these investigations.

Subject to certain modifications the plenary lecture by A. Trautman on the Einstein-Cartan theory also belongs to this framework of ideas. His lecture is concerned with the extension of the Einstein theory by the inclusion of a hypothetical space-time torsion.

The twistor program, the present state of which will be discussed in R. S. Ward's plenary lecture, is also devoted to a unified description of bosons and fermions, admittedly under somewhat different premises.

From these foregoing remarks you will gather that the Einstein idea of a unified field theory is still very much alive, even if we are still unclear which of the approaches adopted so far will lead to a definitive breakthrough.

It is still to be hoped that progress in constructing a unified field theory will also open new prospects for the Mach principle which raises the extremely profound question of the origin of the inertia of mass and the existence of inertial frames of reference. Within the framework of the Einstein theory there is apparently no full solution of Mach's hypothesis.

My personal ambition in the direction of a unified field theory used to lie in the 5-dimensional Projective Theory of Relativity. I began working in this field 25 years ago [4], because I was fascinated by its elegance and compelling logic, particularly after P. Jordan [5] (1945) dropped a constraint hitherto used. Let me also recall in this connection the independent research by A. Einstein and P. G. Bergmann in this field.

I outlined the state of research 15 years ago in the last chapter of my monograph. In the course of the last year I have reanalysed the then occurring difficulties which caused a stagnation in this line of development and I arrived at a basic modification right at the roots. I would like to outline some ideas of this new theory [6, 7].

4. Some remarks on the traditional Projective Relativity Theory

4.1. Arguments for 5-dimensionality

During our congress we obviously shall hear of a lot of new results on quantum gauge theories based on the gauge idea and supersymmetry. Therefore, in the following I will separate the quantum phenomena from my considerations and only dwell at first on Einstein's program of a unified field theory of gravitation and electromagnetism. Of course, I am aware that all new physical theories, concerning their contents of objective reality, must be taken carefully until they have been proved by experiments. Therefore I speak on the following subject with due scientific caution. Why do I advocate taking the 5-dimensionality of physics seriously — more than has been done by the theoreticians during the last two decades? My main arguments are the group equivalence and the intrinsic structure of the field equations.

Group equivalence

The projective approach to gravitation and electromagnetism leads to a prototype of gauge theory in a logically self-contained and aesthetically satisfying way (Latin indices run from 1 to 4, Greek indices run from 1 to 5):

5-dimensional homogeneous coordinate transformations: $X^{\mu'} = X^{\mu'}(X^{\nu})$

→ arbitrary 4-dimensional coordinate transformations: $x^{i'} = x^{i'}(x^j)$

→ gauge transformations: $\tilde{A}_i = A_i + \chi_{,i}, \qquad \tilde{B}_{ij} = B_{ij}$

Field equations

The 5-dimensional field equations yield by 4-bein projection and radial projection into space-time the structure of generalized gravitational equations, generalized electromagnetic equations and a scalar equation:

Notes: a) Jordan's field equations with his exotic parameter $\lambda \gg 1$, introduced by astrophysical arguments are in my opinion not convincing.

b) The above sketched theory has logically nothing to do with the Brans-Dicke ad-hoc theory.

4.2. Difficulties of the conventional scheme

In my opinion the most serious difficulties of the traditional theories (also of my version of 1957) arise from two facts:

a) The generalized gravitational field equation contains second order derivatives of the scalar field, which leads to second order derivatives in the energy tensor of the scalar field. This fact induces physical problems with respect to the positive-definiteness of the energy of the scalar field.

b) The equation of motion of an electrically charged test particle shows a force term proportional to e^2 (e electric charge of the particle). The order of magnitude of this force term is too large and therefore, in my opinion, in contradiction to experience.

Apart from other difficulties both these facts were the main reasons for me to start a new approach from the roots, which will be sketched shortly in the following. Details should be taken from my publications.

5. A new Unified Projective Field Theory

5.1. Fundamentals

In our opinion the most likely approach to a unified field theory — if Einstein's program of geometrization of electromagnetism etc. is reasonable at all — should be based on the following fundamentals:

1. The geometry of the 5-dimensional projective space with curvature and torsion is characterized by the axioms [4]:
   The metric tensor is symmetric.
   The connexion is transvection-invariant, non-symmetric and metric.
2. Projector property of the most important 5-dimensional quantities to guarantee 4-dimensional relationships in space-time after projecting.
3. The 4-dimensional space-time has Riemannian geometry.
4. The use of our vectorial projection formalism [4] for linking both space manifolds is recommended. This formalism allows a short and transparant treatment of these calculations.

It proves that the 5-dimensional geometry based on these fundamentals is so constructed that projection into space-time yields Riemannian geometry for this manifold.
Let us use the following notation (see fig. 1):

X^μ	5-dimensional projective coordinates,
x^i	4-dimensional space-time coordinates,
$\boldsymbol{e}_\mu$	5-dimensional basis vectors (5-bein),
$\boldsymbol{e}_i$	4-dimensional basis vectors (4-bein),
$\boldsymbol{R} = \boldsymbol{e}_\mu X^\mu$	5-dimensional radius vector (projector),
$g_{\mu\nu} = \boldsymbol{e}_\mu \boldsymbol{e}_\nu$	5-dimensional metric,
$g_{ij} = \boldsymbol{e}_i \boldsymbol{e}_j$	4-dimensional metric,
$g_{\mu i} = \boldsymbol{e}_\mu \boldsymbol{e}_i$	mixed metric (projection cosine).

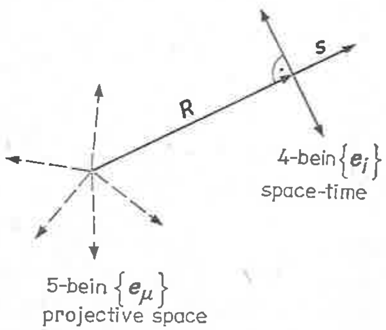


Fig. 1

Apart from the well-known relations

$$\begin{aligned} &\text{a) } \boldsymbol{e}_i = g_{ij}\boldsymbol{e}^j \quad \text{etc.,} && \text{b) } g_{ij}g^{jk} = g_i^k = \delta_i^k, \\ &\text{c) } \boldsymbol{e}_\mu = g_{\mu\nu}\boldsymbol{e}^\nu \quad \text{etc.,} && \text{d) } g_{\mu\nu}g^{\nu\lambda} = g_\mu^\lambda = \delta_\mu^\lambda \end{aligned} \tag{1}$$

one finds the linking relations

$$\begin{aligned} &\text{a) } \boldsymbol{e}_i = g_{i\sigma}\boldsymbol{e}^\sigma \quad \text{etc.,} && \text{b) } g_{i\sigma}g^{j\sigma} = g_i^j, \\ &\text{c) } \boldsymbol{e}_\sigma = g_\sigma^i\boldsymbol{e}_i + s_\sigma\mathbf{s}, && \text{d) } g_i^\mu g_\nu^i = g_\nu^\mu - s_\nu s^\mu, \end{aligned} \tag{2}$$

where

$$\text{a) } s_\nu = \frac{X_\nu}{S}, \qquad \text{b) } \mathbf{s} = \frac{\boldsymbol{R}}{S}, \qquad \text{c) } \boldsymbol{R}^2 = S^2, \qquad \text{d) } \mathbf{s}^2 = 1. \tag{3}$$

For the decomposition of vectors or products yields

$$\begin{aligned} &\text{a) } \boldsymbol{a} = \boldsymbol{e}_\mu a^\mu = \boldsymbol{e}_i a^i + (\boldsymbol{a}\mathbf{s})\,\mathbf{s}, \qquad \text{b) } \boldsymbol{ab} = a_\mu b^\mu = a_i b^i + (\boldsymbol{a}\mathbf{s})\,(\boldsymbol{b}\mathbf{s}), \\ &\text{c) } \chi_\mu\varphi^\mu = \chi_i\varphi^i + (\chi_\mu s^\mu)\,(\varphi_\nu s^\nu) \end{aligned} \tag{4}$$

with

$$\text{a) } a^i = g_\mu^i a^\mu, \qquad \text{b) } \chi_i = g_i^\mu \chi_\mu \quad \text{etc.} \tag{5}$$

With respect to the axiomatics of relating the differential quotient $x^i{}_{|\mu}$ to the mixed metric g_μ^i, important questions in principle arise:

My version of 1957 [4] started in correspondence with Jordan and Ludwig [5] from

$$x^i{}_{|\mu} = g_\mu^i. \tag{6}$$

Some years ago I found it rather interesting to introduce an explicit 4-vector

$$X_i = g_i^\mu X_\mu \tag{7}$$

related to the electromagnetic 4-potential A_i. As to this type of theories I studied two versions [6] in detail:

$$\begin{aligned} &\text{a) } x^i{}_{|\mu} = g_\mu^i - \frac{1}{S^2}\, X^i X_\mu, \\ &\text{b) } x^i{}_{|\mu} = g_\mu^i - \frac{X^i}{S^2 - \Sigma^2}\,(X_\mu - \mathscr{X}_\mu), \end{aligned} \tag{8}$$

where

$$\text{a) } \mathscr{X}_\mu = g_\mu^i X_i, \qquad \text{b) } \Sigma^2 = \mathscr{X}_\mu \mathscr{X}^\mu. \tag{9}$$

Both these variants lead to considerable difficulties with respect to the gauge invariance of the theory. Therefore, I interrupted further work in this direction and devoted my succeeding investigations to the special case $X_i = 0$, but instead of (6) on the basis of the new axiom [7]

$$x^i{}_{|\mu} = e^{\sigma/2} g^i_\mu , \tag{10}$$

where σ ist defined by

$$S = S_0 \, e^\sigma \qquad (S_0 \text{ free constant}). \tag{11}$$

Since we adopt the conventional relation between the 4-dimensional and 5-dimensional coordinates, namely

$$x^i = x^i(X^\mu) \tag{12}$$

to be a homogeneous function of degree 0, the relations

$$\text{a) } x^i{}_{|\mu} X^\mu = 0 \qquad \text{and b) } g^i_\mu \, X^\mu = 0 \tag{13}$$

are valid.
The final version of my theory with the axiom (10) was elaborated in detail after GR 9.

The general covariant derivative (double stroke) and the Riemannian covariant derivative (semicolon) are defined as follows:

$$\text{a) } a^\mu{}_{||\nu} = a^\mu{}_{|\nu} + \Gamma^\mu_{\lambda\nu} a^\lambda \qquad (\Gamma^\mu_{\lambda\nu} \text{ general affinities}),$$

$$\text{b) } a^\mu{}_{;\nu} = a^\mu{}_{|\nu} + \begin{Bmatrix} \mu \\ \lambda\nu \end{Bmatrix} a^\lambda \qquad \left(\begin{Bmatrix} \mu \\ \lambda\nu \end{Bmatrix} \text{ Christoffel symbols} \right). \tag{14}$$

The quantity

$$\sigma_{\mu\nu}{}^\alpha = \Gamma^\alpha_{\mu\nu} - \begin{Bmatrix} \alpha \\ \mu\nu \end{Bmatrix} \tag{15}$$

is related to torsion.

As in the conventional theory we use the 5-dimensional curl tensor

$$X_{\mu\nu} = X_{\nu|\mu} - X_{\mu|\nu} \tag{16}$$

whose 4-bein projection reads

$$X_{mn} = g^\mu_m g^\nu_n X_{\mu\nu} . \tag{17}$$

In this connection one should mention the validity of the Killing equation

$$X_{\mu;\nu} + X_{\nu;\mu} = 0 \tag{18}$$

for the quantities X_μ.

For some practical reasons of calculation it is convenient to introduce the curl

$$s_{\mu\nu} = s_{\nu|\mu} - s_{\mu|\nu} = \frac{1}{S} X_{\mu\nu} + s_\mu \sigma_{|\nu} - s_\nu \sigma_{|\mu} \tag{19}$$

whose 4-bein projection is

$$s_{mn} = g^\mu_m g^\nu_n s_{\mu\nu} = \frac{1}{S} X_{mn} . \tag{20}$$

Further

$$s_{\mu\nu}s^{\nu} = \sigma_{|\mu} \tag{21}$$

holds.

On our geometrical axiomatics the constancy of the unit vector (3a, b) can be proved:

$$\text{a) } s_{\nu||\mu} = 0 \quad \text{or} \quad \text{b) } \boldsymbol{s}_{|\mu} = 0 \quad \text{or} \quad \text{c) } \boldsymbol{s}_{|i} = 0 \tag{22}$$

(already found in our former papers, see e.g. our monograph [4]).

Furthermore, for the torsion vector the explicit expression

$$\boldsymbol{H}_{\nu\mu} = \boldsymbol{e}_{\nu|\mu} - \boldsymbol{e}_{\mu|\nu} = \boldsymbol{s}\left[s_{\mu\nu} + \frac{1}{2}(s_\nu\sigma_{|\mu} - s_\mu\sigma_{|\nu})\right] + \frac{1}{2}(\boldsymbol{e}_\mu\sigma_{|\nu} - \boldsymbol{e}_\nu\sigma_{|\mu}) \tag{23}$$

results, whereas for the torsion tensor

$$S_{\nu\mu}{}^{\tau} = \frac{1}{2}(\Gamma^{\tau}_{\nu\mu} - \Gamma^{\tau}_{\mu\nu}) \tag{24}$$

the relation

$$2S_{\nu\mu\tau} = s_\tau s_{\mu\nu} + \frac{1}{2}(g_{\mu\tau} - s_\mu s_\tau)\,\sigma_{|\nu} - \frac{1}{2}(g_{\nu\tau} - s_\nu s_\tau)\,\sigma_{|\mu} \tag{25}$$

holds. For the combination

$$\sigma_{\lambda\mu\varkappa} = S_{\lambda\varkappa\mu} + S_{\lambda\mu\varkappa} + S_{\mu\varkappa\lambda}, \tag{26}$$

defined by (15), we find

$$2\sigma_{\nu\tau\mu} = s_\tau s_{\mu\nu} + s_\mu s_{\tau\nu} + s_\nu s_{\mu\tau} + (g_{\mu\tau} - s_\mu s_\tau)\,\sigma_{|\nu} - (g_{\nu\tau} - s_\nu s_\tau)\,\sigma_{|\mu}. \tag{27}$$

5.2. Projection theorems for derivatives

On the basis of the new postulate (10) the projection theorems take the shape:

$$\text{a) } I_{|\nu}g^{\nu}_{n} = e^{\sigma/2}I_{|n}, \qquad \text{b) } a_{\nu||\mu}g^{\nu}_{n}g^{\mu}_{m} = e^{\sigma/2}a_{n;m}, \tag{28}$$

$$\text{c) } a_{\nu\sigma||\mu}g^{\nu}_{n}g^{\sigma}_{s}g^{\mu}_{m} = e^{\sigma/2}a_{ns;m},$$

$$\text{d) } a_{\sigma||\nu||\mu}g^{\sigma}_{s}g^{\nu}_{n}g^{\mu}_{m} = e^{\sigma}\left[a_{s;n;m} + \frac{1}{2}\sigma_{|m}a_{s;n}\right].$$

For a general vector $\boldsymbol{a}$ the equation

$$\boldsymbol{a}_{||\nu||\mu}g^{\nu}_{n}g^{\mu}_{m} = e^{\sigma}\left[\boldsymbol{a}_{|n;m} + \frac{1}{2}\sigma_{|m}\boldsymbol{a}_{|n}\right] \tag{29}$$

yields, whereas for the basis vectors

$$\boldsymbol{e}_{\nu||\mu}g^{\nu}_{n}g^{\mu}_{m} = e^{\sigma/2}\boldsymbol{e}_{n;m} \tag{30}$$

is valid. This relation shows the consistency of the geometrical axioms: $\boldsymbol{e}_{n;m} = 0$ and $\boldsymbol{e}_{\nu||\mu} = 0$ (metricity). For practical calculation the formula

$$F_{|\varkappa;\mu} g^{\varkappa}_{k} g^{\mu}_{m} = \mathrm{e}^{\sigma} \left[F_{|k;m} + \frac{1}{2} \left(F_{|k}\sigma_{|m} + F_{|m}\sigma_{|k} - g_{km} F_{|j}\sigma^{|j} \right) \right] \tag{31}$$

is useful.

5.3. Cyclic relations and curvature tensors

With respect to the cyclic Maxwell system of electromagnetism the existence of a 4-dimensional cyclic relation is of interest:

The 5-dimensional cyclic relations

$$\text{a)}\ X_{\langle\alpha\varrho|\tau\rangle} = 0, \qquad \text{b)}\ s_{\langle\alpha\varrho|\tau\rangle} = 0 \tag{32}$$

follow immediately from the definition of $X_{\alpha\varrho}$ and $s_{\alpha\varrho}$ as curls. Equivalent to (32) are the equations

$$\text{a)}\ \varepsilon_{\mu\nu}{}^{\tau\varrho\sigma} X_{\varrho\sigma;\tau} = 0, \qquad \text{b)}\ \varepsilon_{\mu\nu}{}^{\tau\varrho\sigma} s_{\varrho\sigma;\tau} = 0 \tag{33}$$

($\varepsilon_{\mu\nu}{}^{\tau\varrho\sigma}$ 5-dimensional Levi-Civita pseudotensor). Using the relation [4]

$$\varepsilon_m{}^{nks} = \overset{5}{s_{\nu}} \varepsilon_m{}^{nks\nu} \tag{34}$$

between the 5-dimensional and the 4-dimensional Levi-Civita pseudotensors and the projection theorems, we arrive at the 4-dimensional cyclic relations

$$\text{a)}\ \varepsilon^{mjkl} \left(\frac{X_{kl}}{S^3} \right)_{;j} = 0 \quad \text{resp.} \quad \text{b)}\ \left[\left(\frac{X_{kl}}{S^3} \right)_{;j} \right]_{\langle klj\rangle} = 0. \tag{35}$$

The 5-dimensional general curvature tensor $G^{\nu}{}_{\varkappa\mu\lambda}$ is defined by

$$T_{\varkappa||\mu||\lambda} - T_{\varkappa||\lambda||\mu} = T_{\varkappa|\mu|\lambda} - T_{\varkappa|\lambda|\mu} + T_{\nu} G^{\nu}{}_{\varkappa\mu\lambda} + 2T_{\varkappa||\alpha} S_{\lambda\mu}{}^{\alpha}. \tag{36}$$

Expressed by the affinities it reads

$$G^{\nu}{}_{\varkappa\mu\lambda} = \Gamma^{\nu}_{\varkappa\lambda|\mu} - \Gamma^{\nu}_{\varkappa\mu|\lambda} + \Gamma^{\tau}_{\varkappa\lambda}\Gamma^{\nu}_{\tau\mu} - \Gamma^{\tau}_{\varkappa\mu}\Gamma^{\nu}_{\tau\lambda}. \tag{37}$$

It has the symmetry properties

$$\text{a)}\ G_{\sigma\varkappa\mu\lambda} = -G_{\varkappa\sigma\mu\lambda}, \qquad \text{b)}\ G_{\sigma\varkappa\mu\lambda} = -G_{\sigma\varkappa\lambda\mu}. \tag{38}$$

Furthermore, the generalized cyclic relation

$$G_{\alpha\langle\nu\tau\sigma\rangle} = 2[S_{\nu\sigma\alpha||\tau} - 2S_{\sigma\tau}{}^{\beta} S_{\beta\nu\alpha}]_{\langle\nu\sigma\tau\rangle} \tag{39}$$

and the generalized Bianchi identity

$$[G_{\beta\mu\sigma\nu||\tau} - 2G_{\beta\mu\nu\alpha} S_{\sigma\tau}{}^{\alpha}]_{\langle\nu\sigma\tau\rangle} \doteq 0 \tag{40}$$

are valid, from which the contracted Bianchi identity

$$\left(G^{\mu}{}_{\tau} - \frac{1}{2} g^{\mu}_{\tau} \overset{5}{G}\right)_{||\mu} = 2G^{\mu}{}_{\alpha} S_{\mu\tau}{}^{\alpha} - G^{\beta\mu}{}_{\tau\alpha} S_{\beta\mu}{}^{\alpha} \tag{41}$$

results, where

$$\text{a) } G_{\mu\tau} = G^{\beta}{}_{\mu\tau\beta}, \qquad \text{b) } \overset{5}{G} = G^{\beta}{}_{\beta}. \tag{42}$$

The analogous formulas for the Riemannian geometry can be obtained by the specialization $G \to R$.

Let us conclude this section by reproducing some important formulas resulting from the projector postulates:

$$\begin{aligned} &\text{a) } G_{\mu\sigma\tau\lambda} X^{\lambda} = 0, \\ &\text{b) } X_{\nu} R^{\nu}{}_{\varkappa\mu\lambda} = \frac{1}{2} X_{\mu\lambda;\varkappa}, \\ &\text{c) } X^{\nu} X^{\lambda} R_{\nu\varkappa\mu\lambda} = S_{|\mu} S_{|\varkappa} + S S_{|\varkappa;\mu} + \frac{1}{4} X_{\varkappa\lambda} X^{\lambda}{}_{\mu}. \end{aligned} \tag{43}$$

5.4. Decomposition formulas for the curvature tensor

Using the algebraic relations previously derived, we arrive at the following auxiliary formulas:

$$\text{a) } S_{|\mu} S^{|\mu} = e^{\sigma} S_{|j} S^{|j}, \qquad \text{b) } S^{|\mu}{}_{;\mu} = e^{\sigma} S^{|j}{}_{;j}, \tag{44}$$

$$\begin{aligned} &\text{a) } \overset{5}{X}_{n\lambda} \overset{5}{X}_{m}{}^{\lambda} = X_{nj} X_{m}{}^{j} + 4 e^{\sigma} S_{|n} S_{|m}, \\ &\text{b) } X_{\nu\lambda} X^{\nu\lambda} = X_{kj} X^{kj} + 8 e^{\sigma} S_{|j} S^{|j}, \\ &\text{c) } g^{\lambda}_{l} X^{\mu}{}_{\lambda||\mu} = e^{\sigma/2} X^{j}{}_{l;j}. \end{aligned} \tag{45}$$

Applying this decomposition procedure to the curvature and torsion quantities, we find

$$\begin{aligned} &\text{a) } \overset{5}{R}_{nl} = g^{\nu}_{n} g^{\lambda}_{l} R_{\nu\lambda} = \frac{1}{2} s_{nk} s_{l}{}^{k} + e^{\sigma} \left[\overset{4}{R}_{nl} + \frac{3}{2} \sigma_{|n} \sigma_{|l} - \frac{1}{2} g_{nl} \sigma^{|k}{}_{;k}\right], \\ &\text{b) } \overset{5}{R} = \frac{1}{4} s_{nk} s^{nk} + e^{\sigma} \left(\overset{4}{R} + \frac{3}{2} \sigma_{|k} \sigma^{|k} - \sigma^{|k}{}_{;k}\right), \\ &\text{c) } X^{\nu} \overset{5}{R}_{\nu l} = \frac{1}{2} e^{\sigma/2} X^{j}{}_{l;j}, \\ &\text{d) } X^{\nu} X^{\lambda} R_{\nu\lambda} = -\frac{1}{4} X_{kn} X^{kn} + S^{2} e^{\sigma} \sigma^{|n}{}_{;n}. \end{aligned} \tag{46}$$

In deriving these decomposition formulas the relations

$$X^\nu G_{\nu\varkappa\mu\lambda} = 0 \tag{47}$$

and

$$G_{skml} = g^\sigma_s g^\varkappa_k g^\mu_m g^\lambda_l G_{\sigma\varkappa\mu\lambda} = \mathrm{e}^\sigma \overset{4}{R}_{skml} \tag{48}$$

were used.

Though the 5-dimensional coordinate differential $\mathrm{d}X^\mu$ is not a projector itself, fortunately its projection according to (10),

$$g^i_\mu \, \mathrm{d}X^\mu = \mathrm{e}^{-\sigma/2} \, \mathrm{d}x^i, \tag{49}$$

leads to a 4-dimensional quantity. This means that (with some caution) we can use the 5-dimensional coordinate differential and the tangential vector as basic concepts of our theory.

5.5. Field equations and conservation laws in the 5-dimensional projective space

As we already did in 1957, we start from the 5-dimensional field equation

$$R^{\mu\tau} - \frac{1}{2} g^{\mu\tau} \overset{5}{R} + D^{\mu\tau} = \varkappa_0 \theta^{\mu\tau} \qquad (\varkappa_0 \text{ free constant}), \tag{50}$$

where $\theta^{\mu\tau}$ is the "substrate projector" of the non-geometrized matter, and

$$D^{\mu\tau} = \lambda_0 S^C (g^{\mu\tau} + C s^\mu s^\tau) \qquad (\lambda_0, C \text{ free constants}) \tag{51}$$

with

$$D^{\mu\tau}{}_{;\tau} = 0 \tag{52}$$

means a generalization of the so-called cosmological term. Then the conservation law

$$\theta^{\mu\tau}{}_{;\tau} = 0 \tag{53}$$

follows. Taking into account that from (50) also the relation

$$\overset{5}{R} = R^\mu{}_\mu = \frac{2}{3} \lambda_0 S^C (C + 5) - \frac{2}{3} \varkappa_0 \overset{5}{\theta} \tag{54}$$

with

$$\overset{5}{\theta} = \theta^\mu{}_\mu \tag{55}$$

results, we can give the field equation (50) the different form

$$R^{\mu\tau} - \frac{1}{3} (C + 2) \lambda_0 S^C g^{\mu\tau} - \lambda_0 C S^C s^\mu s^\tau = \varkappa_0 \left(\theta^{\mu\tau} - \frac{1}{3} \overset{5}{\theta} g^{\mu\tau}\right). \tag{56}$$

5.6. Field equations and conservation laws expressed by 4-dimensional geometrical quantities

5.6.1. Tensorial field equation

Using the abbreviations

$$\text{a)}\ \overset{5}{\theta}{}^{mn} = \theta^{\mu\tau} g^m_\mu g^n_\tau, \qquad \text{b)}\ \overset{5}{\theta}{}^m{}_m = \overset{5}{\theta}{}^{mn} g_{mn} \tag{57}$$

and the relations

$$\text{a)}\ \overset{5}{R} = \overset{5}{R}{}^m{}_m + R_{\mu\nu} s^\mu s^\nu, \qquad \text{b)}\ \overset{5}{\theta} = \overset{5}{\theta}{}^m{}_m + \theta_{\mu\nu} s^\mu s^\nu, \tag{58}$$

we find by 4-bein projection of (50) the tensor equation

$$\text{a)}\ \overset{5}{R}_{mn} - \frac{1}{2} g_{mn} \overset{5}{R} + \lambda_0 S^C g_{mn} = \varkappa_0 \overset{5}{\theta}_{mn} \text{ resp.}$$

$$\text{b)}\ \overset{5}{R}_{mn} = \frac{\lambda_0}{3} (C + 2)\, S^C g_{mn} + \varkappa_0 \left(\overset{5}{\theta}_{mn} - \frac{1}{3} g_{mn} \overset{5}{\theta} \right) \tag{59}$$

and hence

$$\overset{5}{R} = \overset{5}{R}{}^m{}_m + \varkappa_0 \left(\frac{2}{3} \overset{5}{\theta} - \overset{5}{\theta}{}^m{}_m \right) + \frac{2\lambda_0}{3} (1 - C)\, S^C. \tag{60}$$

Rearranging (59) by means of (46a) and (46b) gives the 4-dimensional tensorial field equation

$$\overset{4}{R}_{mn} - \frac{1}{2} g_{mn} \overset{4}{R} + \lambda_0 S^C \,\mathrm{e}^{-\sigma} g_{mn} = \varkappa_0 \,\mathrm{e}^{-\sigma} \overset{5}{\theta}_{mn} + \frac{1}{2} \mathrm{e}^{-\sigma} \left[s_{mk} s^k{}_n + \frac{1}{4} g_{mn} s_{jk} s^{jk} \right]$$
$$- \frac{3}{2} \left[\sigma_{|m} \sigma_{|n} - \frac{1}{2} g_{mn} \sigma_{|k} \sigma^{|k} \right]. \tag{61}$$

5.6.2. Vectorial field equation

Radial projection of (50) leads to the vector equation

$$R^{\mu\tau} X_\tau = \varkappa_0 \theta^{\mu\tau} X_\tau - \frac{1}{3} X^\mu \left[\varkappa_0 \overset{5}{\theta} + 2\lambda_0 (C - 1)\, S^C \right]. \tag{62}$$

Further 4-bein projection yields

$$\overset{5}{R}{}^{m\tau} X_\tau = \varkappa_0 \overset{5}{\theta}{}^{m\tau} X_\tau \tag{63}$$

if we introduce the abbreviation

$$\overset{5}{\theta}{}^{m\tau} = \theta^{\mu\tau} g^m_\mu . \tag{64}$$

Using the results (46c), instead of (63) we arrive at the vectorial field equation

$$X^{mj}{}_{;j} = -2\varkappa_0 \,\mathrm{e}^{-\sigma/2} \overset{5}{\theta}{}^m{}_\tau X^\tau . \tag{65}$$

5.6.3. Scalar field equation

Our next step is to perform a further radial projection of (62). By means of (58b) we find

$$R^{\mu\tau} X_\mu X_\tau = \varkappa_0 S^2 \left(\frac{2}{3}\,\overset{5}{\theta} - \overset{5}{\theta}{}^{m}{}_{m}\right) - \frac{2}{3}\,\lambda_0 (C-1)\, S^{C+2}. \tag{66}$$

Rearranging (66) with the help of (46d) gives the scalar field equation

$$\sigma^{|k}{}_{;k} = \varkappa_0\, \mathrm{e}^{-\sigma} \left(\frac{2}{3}\,\overset{5}{\theta} - \overset{5}{\theta}{}^{m}{}_{m}\right) + \frac{1}{4}\, \mathrm{e}^{-\sigma} s_{kj} s^{kj} - \frac{2}{3}\, \lambda_0 (C-1)\, \mathrm{e}^{-\sigma} S^C. \tag{67}$$

5.6.4. Conservation laws

For the purpose of deriving the conservation laws we first rewrite (53) as

$$\theta^{\mu\tau}{}_{||\tau} = \theta^{\lambda\tau} \sigma_{\lambda\tau}{}^{\mu} + \theta^{\mu\lambda} \sigma_{\lambda\tau}{}^{\tau}. \tag{68}$$

Eliminating the torsion quantities by means of (27) and performing 4-bein projection leads to the vectorial conservation law

$$\left(\frac{1}{S}\, \overset{5}{\theta}{}^{jm}\right)_{;m} = \frac{1}{S}\, \mathrm{e}^{-\sigma/2} s^{j}{}_{m} \overset{5}{\theta}{}^{m}{}_{\tau} s^{\tau} + \frac{1}{S} \left(\overset{5}{\theta} - \frac{3}{2}\, \overset{5}{\theta}{}^{m}{}_{m}\right) \sigma^{|j}, \tag{69}$$

whereas radial projection gives the continuity equation

$$\left(\mathrm{e}^{\sigma/2} \overset{5}{\theta}{}^{m\mu} s_\mu\right)_{;m} = 0. \tag{70}$$

5.7. Evaluation of the field equations and conservation laws

5.7.1. Cyclic Maxwell system of electromagnetism

It is obvious that the cyclic relation (35b) offers to be interpreted as the cyclic (homogeneous) Maxwell system of electromagnetism

$$B_{\langle kl;j\rangle} = 0 \tag{71}$$

if we identify the electromagnetic field strength tensor as

$$B_{kl} = \frac{e_0 S_0}{S^3}\, X_{kl} = \frac{e_0 S_0}{S^2}\, s_{kl}, \tag{72}$$

where according to dimensional considerations e_0 may be the universal constant "electric elementary charge".

5.7.2. Inhomogeneous Maxwell system of electromagnetism

If we identify the electromagnetic induction tensor as

$$H^{mj} = b_0 X^{mj} = b_0 S s^{mj} \qquad (b_0 \text{ free constant}) \tag{73}$$

and if we introduce the electric current density by

$$j^m = -\frac{c\varkappa_0 b_0 S_0}{2\pi}\, e^{\sigma/2} \overset{5}{\theta}{}^m{}_\tau s^\tau, \tag{74}$$

the vectorial field equation (65) takes the shape of the inhomogeneous Maxwell system (CGS-system of units):

$$H^{mj}{}_{;j} = \frac{4\pi}{c}\, j^m. \tag{75}$$

I would like to introduce for the phenomenon connected with the scalar field S (dimension of lenght) resp. σ (dimensionless) the notion "scalarism" in analogy to electromagnetism. If such a new phenomenon exists in nature, a prediction which, of course, can only be decided by experiments, this would mean that even in the case of vacuum a scalaristic polarisation phenomenon according to the relation

$$H_{kl} = \bar{\varepsilon} B_{kl}, \tag{76}$$

where

$$\bar{\varepsilon} = \frac{b_0}{e_0 S_0}\, S^3 \tag{77}$$

is the scalaristic dielectricity of the vacuum, has to be expected (see also [8]).

5.7.3. Field equation of scalarism

Considering the scalar field equation (67), we form the opinion that the true physical substrate quantities should be defined by

$$\text{a)}\ \theta_{mn} = \frac{S_0}{S} \overset{5}{\theta}_{mn} = e^{-\sigma} \overset{5}{\theta}_{mn} \qquad \text{(energy tensor of the substrate)},$$

$$\text{b)}\ \theta = \frac{S_0}{S} \overset{5}{\theta} = e^{-\sigma} \overset{5}{\theta} \qquad \text{(trace)}. \tag{78}$$

Then by means of (72) and (73) the equation (67) goes over into the field equation of scalarism

$$\sigma^{|k}{}_{;k} = \varkappa_0 \left(\frac{2}{3}\,\theta - \theta^m{}_m\right) + \frac{1}{4e_0 b_0}\, B_{kj} H^{kj} - \frac{2\lambda_0 S_0}{3}\,(C-1)\, S^{C-1}. \tag{79}$$

5.7.4. Generalized field equation of gravitation

Let us now propose a reasonable physical interpretation of the tensorial field equation (61). First we introduce the electromagnetic energy tensor (Minkowski tensor)

$$E_{mn} = \frac{1}{4\pi}\left(B_{mk} H^k{}_n + \frac{1}{4}\, g_{mn} B_{jk} H^{jk}\right) \tag{80}$$

and the scalaristic energy tensor

$$\Sigma_{mn} = -\frac{3}{2\varkappa_0}\left(\sigma_{|m}\sigma_{|n} - \frac{1}{2}\, g_{mn}\sigma_{|k}\sigma^{|k}\right). \tag{81}$$

Then the generalized gravitational field equation takes the form

$$\overset{4}{R}_{mn} - \frac{1}{2}\, g_{mn}\overset{4}{R} + \lambda_0 S_0 S^{C-1} g_{mn} = \varkappa_0(\theta_{mn} + E_{mn} + \Sigma_{mn}). \tag{82}$$

Conclusion:

The free constant $\varkappa_0$ has to be identified with Einstein's gravitational constant. This means that the gravitational coupling factor remains a true constant, in contrast to all predictions concerning a variability of the gravitational constant.

Furthermore, we come to the result that scalarism induces gravitation by means of the energy tensor (81). This scalaristic energy tensor consists of partial derivatives of first order and therefore avoids the physical difficulties of the previous theories which were in the dilemma of having an energy tensor with second order derivatives.

From (81) for the scalaristic energy density the expression

$$\Sigma_4{}^4 = \frac{3}{4\varkappa_0}\left(\sum_{a=1}^{3}\sigma_{|a}\sigma^{|a} - \sigma_{|4}\sigma^{|4}\right) \tag{83}$$

results, which specializes for a static field to the positive definite quantity

$$(\Sigma_4{}^4)_{\text{stat}} = \frac{3}{4\varkappa_0}\sum_{a=1}^{3}\sigma_{|a}\sigma^{|a} \geqq 0. \tag{84}$$

5.7.5. Equation of motion of the substrate

We identify the vectorial conservation law (69) as the equation of motion of the substrate. Indeed, by means of (72), (74) and (78) we obtain the structure of the equation of motion of a continuum, namely

$$\theta^{jm}{}_{;m} = -\frac{1}{c}\, B^j{}_m j^m + \left(\theta - \frac{3}{2}\,\theta^m{}_m\right)\sigma^{|j} \tag{85}$$

if we fix the free constant b_0 by

$$b_0 = \frac{2\pi}{e_0\varkappa_0}. \tag{86}$$

The Lorentz force density

$$^{(L)}f^j = -\frac{1}{c}\, B^j{}_m j^m \tag{87}$$

appears in a quite natural way automatically. But furthermore the scalaristic force density

$$^{(S)}f^j = \left(\theta - \frac{3}{2}\,\theta^m{}_m\right)\sigma^{|j} \tag{88}$$

is predicted.

5.7.6. Continuity equation

Considering (70) and (74) we immediately find out that (70) has to be identified as the continuity equation for the electric current density, i.e.

$$j^m{}_{;m} = 0\,. \tag{89}$$

Thus the equation of motion and the conservation of the electric charge prove to be consequences of the field equations.

5.7.7. Discussion of the free constants

The free constants $\varkappa_0$ and b_0 have been already fixed above, i.e. we are left with the further free constants S_0, λ_0 and C, where S_0 has the physical dimension of length. Let us now in our next step eliminate b_0 in (77) by means of (86):

$$\bar{\varepsilon} = \frac{2\pi}{e_0{}^2 S_0 \varkappa_0}\,S^3\,. \tag{90}$$

It seems to be reasonable to demand that in the "infinitely diluted state of matter" (infinity of an isolated system)

$$\text{a) } S|_\infty = S_0, \quad \text{i.e.} \qquad \text{b) } \sigma|_\infty = 0 \tag{91}$$

and

$$\text{a) } B_{kj}|_\infty = H_{kj}|_\infty\,, \quad \text{i.e.} \qquad \text{b) } \bar{\varepsilon}|_\infty = 1\,. \tag{92}$$

Applying these boundary conditions to (90), we find for the scalaristic dielectricity of the vacuum

$$\bar{\varepsilon} = \left(\frac{S}{S_0}\right)^3 = \mathrm{e}^{3\sigma} \tag{93}$$

and further the interesting equation for the constant S_0:

$$S_0 = e_0 \sqrt{\frac{\varkappa_0}{2\pi}} \tag{94}$$

which plays the role of a fundamental constant of the physical dimension of length. We propose to call it the universal constant "scalaristic elementary length". Remembering the numerical values for Einstein's gravitational constant $\varkappa_0$ and the electric elementary charge e_0 (γ_N Newton's gravitational constant):

$$\varkappa_0 = \frac{8\pi\gamma_N}{c^4} = 2.08 \cdot 10^{-48}\,\mathrm{g^{-1}\,cm^{-1}\,s^2}, \quad e_0 = 4.8 \cdot 10^{-10}\,\mathrm{g^{1/2}\,cm^{3/2}\,s^{-1}}, \tag{95}$$

we find [4]:

$$S_0 = 2.76 \cdot 10^{-34}\,\mathrm{cm}. \tag{96}$$

Eliminating the electric elementary charge e_0 in (94) in favour of Sommerfeld's fine structure constant

$$\alpha_S = \frac{e_0{}^2}{\hbar c} = \frac{1}{137.04}, \tag{97}$$

we get

$$S_0 = \sqrt{\frac{\varkappa_0 \hbar c \alpha_S}{2\pi}}. \tag{98}$$

Usually Planck's elementary length is defined by

$$l_P = \sqrt{\frac{\hbar\gamma_N}{c^3}} = \frac{1}{2}\sqrt{\hbar\varkappa_0 c} = 4.05 \cdot 10^{-33}\,\mathrm{cm}. \tag{99}$$

Comparing it with (98) leads to the relation:

$$S_0 = l_P \sqrt{\frac{2\alpha_S}{\pi}}. \tag{100}$$

Our further considerations concentrate on the scalaristic field equation (79). It seems reasonable to demand according to the boundary conditions (91) and (92) for isolated systems

$$\sigma^{|k}{}_{;k}|_\infty = 0. \tag{101}$$

This condition means

$$C = 1. \tag{102}$$

Introducing the cosmological constant

$$\lambda_C = \lambda_0 S_0 \tag{103}$$

whose justification is on the same footing as in Einstein's theory, from (82) and (79) we finally get the generalized gravitational field equation

$$\overset{4}{R}_{mn} - \frac{1}{2}\, g_{mn}\overset{4}{R} + \lambda_C g_{mn} = \varkappa_0(\Theta_{mn} + E_{mn} + \Sigma_{mn}) \tag{104}$$

and the scalaristic field equation

$$\sigma^{|k}{}_{;k} = \frac{2}{3}\,\varkappa_0\left(\Theta - \frac{3}{2}\,\Theta^m{}_m\right) + \frac{\varkappa_0}{8\pi}\, B_{kj}H^{kj}. \tag{105}$$

Conclusion:

Since according to our theory electromagnetism induces as a source term scalarism (105), and scalarism via the scalaristic energy tensor Σ_{mn} as a source term gravitation (104), our theory does not contain the Einstein-Maxwell theory (simple superposition of gravitation and electromagnetism) as the special case $\sigma = 0$. This is a question of principle: If the Einstein-Maxwell theory is valid, our theory must be wrong and vice versa. Of course, this basic question can only be decided by an experimentum crucis.

Let us finally note that from (80) and (81) the useful force density relations

$$E_m{}^n{}_{;n} = \frac{1}{c} B_{mn} j^n + \frac{3}{16\pi} B_{jk} H^{jk} \sigma_{|m} \tag{106}$$

and

$$\Sigma_m{}^n{}_{;n} = -\left(\theta - \frac{3}{2} \theta^n{}_n\right) \sigma_{|m} - \frac{3}{16\pi} B_{jk} H^{jk} \sigma_{|m} \tag{107}$$

result.

5.8. Coordinate differential and line element

The 5-dimensional and 4-dimensional line element vectors

$$\text{a)}\ \mathrm{d}\overset{5}{\boldsymbol{s}} = \boldsymbol{e}_\mu \,\mathrm{d}X^\mu, \qquad \text{b)}\ \mathrm{d}\overset{4}{\boldsymbol{s}} = \boldsymbol{e}_i \,\mathrm{d}x^i \tag{108}$$

resp. the squares of the line elements

$$\text{a)}\ \left(\mathrm{d}\overset{5}{s}\right)^2 = g_{\mu\nu} \,\mathrm{d}X^\mu \,\mathrm{d}X^\nu, \qquad \text{b)}\ \left(\mathrm{d}\overset{4}{s}\right)^2 = g_{ij} \,\mathrm{d}x^i \,\mathrm{d}x^j \tag{109}$$

obey the relations

$$\text{a)}\ \mathrm{d}\overset{5}{\boldsymbol{s}} = \mathrm{e}^{-\sigma/2} \,\mathrm{d}\overset{4}{\boldsymbol{s}} + \boldsymbol{s}\beta \,\mathrm{d}\overset{5}{s}, \qquad \text{b)}\ \left(\mathrm{d}\overset{5}{s}\right)^2 = \frac{\mathrm{e}^{-\sigma}}{1-\beta^2} \left(\mathrm{d}\overset{4}{s}\right)^2, \tag{110}$$

where the abbreviation

$$\beta = s_\nu t^\nu = \boldsymbol{s}\boldsymbol{t} \tag{111}$$

with

$$\text{a)}\ t^\nu = \frac{\mathrm{d}X^\nu}{\mathrm{d}\overset{5}{s}}, \qquad \text{b)}\ \boldsymbol{t} = \boldsymbol{e}_\nu t^\nu \qquad \text{(tangential vector)} \tag{112}$$

was used.

5.9. Metric projection tensor

As it is well known, from the 4-dimensional metric tensor g^{ij} and the 4-velocity $u^i = \frac{dx^i}{d\tau}$ the 4-dimensional metric projection tensor is constructed as follows:

$$h^{ij} = g^{ij} + \frac{1}{c^2} u^i u^j. \tag{113}$$

For this quantity the relations

$$\text{a) } h_i{}^k h_{mk} = h_{im}, \qquad \text{b) } h_{ik} u^k = 0, \qquad \text{c) } h^m{}_m = 3, \qquad \text{d) } h_{ik} h^{ik} = 3 \tag{114}$$

hold.

Let us now introduce the 5-velocity by

$$u^\mu = \mathrm{i}c \frac{dX^\mu}{d\overset{5}{s}} = \mathrm{i}ct^\mu \tag{115}$$

with

$$u_\mu u^\mu = -c^2. \tag{116}$$

By means of (112) we find

$$u_\mu s^\mu = \mathrm{i}c\beta. \tag{117}$$

Furthermore, because of (110)

$$\overset{5}{u}{}^i = g^i_\mu u^\mu = \sqrt{1-\beta^2}\, u^i \tag{118}$$

is valid.

For several reasons it seems to be convenient to define the 5-dimensional metric projection tensor by*)

$$h^{\mu\nu} = g^{\mu\nu} + \frac{1}{c^2} u^\mu u^\nu. \tag{119}$$

From this definition the following relations result:

$$\begin{gathered} \text{a) } h^{\mu\nu} u_\nu = 0, \qquad \text{b) } h^{\mu\nu} s_\nu = s^\mu + \frac{\mathrm{i}\beta}{c} u^\mu, \\ \text{c) } h^{\mu\nu} s_\mu s_\nu = 1-\beta^2, \qquad \text{d) } h^{\mu\nu} h_{\mu\lambda} = h^\nu{}_\lambda, \\ \text{e) } h^\mu{}_\mu = 4, \qquad \text{f) } h_{\mu\nu} h^{\mu\nu} = 4, \qquad \text{g) } h^{\mu\nu} g^i_\mu g^j_\nu = h^{ij} - \frac{\beta^2}{c^2} u^i u^j. \end{gathered} \tag{120}$$

*) See for a different definition Erice proceedings [7].

5.10. Equation of motion for a test particle

5.10.1. Investigation of a modified geodesic

Let us first investigate the following variational principle in the projective space as a mathematical problem:

$$\delta \int_{P_0}^{P_1} f(S)\, \mathrm{d}\overset{5}{s} = 0 \tag{121}$$

with

$$\delta X^\mu|_{P_1} = \delta X^\mu|_{P_0} = 0\,. \tag{122}$$

Using the well known relation

$$\delta\left(\mathrm{d}\overset{5}{s}\right) = \frac{1}{2}\, g_{\mu\nu|\alpha} t^\mu t^\nu \delta X^\alpha\, \mathrm{d}\overset{5}{s} + g_{\mu\alpha} t^\mu\, \mathrm{d}(\delta X^\alpha) \tag{123}$$

and remembering (112), we obtain as a necessary condition for satisfying (121) the following differential equation

$$t_{\alpha;\beta} t^\beta = \frac{f'}{f}\left(S_{|\alpha} - \frac{\mathrm{d}S}{\mathrm{d}\overset{5}{s}}\, t_\alpha\right) \tag{124}$$

(prime means differentiation with respect to S). Radial projection leads to

$$\beta = \frac{\beta_0}{fS} \qquad (\beta_0 \text{ constant}), \tag{125}$$

whereas 4-bein projection yields

$$\frac{\mathrm{D}\left(u^i\sqrt{1-\beta^2}\right)}{\mathrm{D}\tau} = \left(u^i\sqrt{1-\beta^2}\right)_{;j} u^j = \frac{\mathrm{i}c\beta_0 S_0^{1/2}}{fS^{5/2}}\, X^i{}_{;j} u^j$$
$$- \frac{c^2}{\sqrt{1-\beta^2}}\left(\frac{3}{2}\beta^2 - \frac{1}{2} + \frac{f'S}{f}\right)\sigma^{|i} + \frac{\mathrm{d}\sigma}{\mathrm{d}\tau}\left(\frac{1}{2} - \frac{f'S}{f}\right)\sqrt{1-\beta^2}\, u^i \tag{126}$$

or by means of (72)

$$\frac{\mathrm{D}u^i}{\mathrm{D}\tau} = \frac{\mathrm{i}c\beta_0\, \mathrm{e}^{\sigma/2}}{fe_0\sqrt{1-\beta^2}}\, B^i{}_{;j} u^j + \frac{c^2}{1-\beta^2}\left(\frac{1}{2} - \frac{3}{2}\beta^2 - \frac{f'S}{f}\right)\sigma^{|i}$$
$$+ \frac{1}{1-\beta^2}\frac{\mathrm{d}\sigma}{\mathrm{d}\tau}\left(\frac{1}{2} - \frac{3}{2}\beta^2 - \frac{f'S}{f}\right)u^i. \tag{127}$$

5.10.2. Equation of motion

The decision on the choice of $f(S)$ is now rather complicated. We studied different possibilities:

(i) Reflecting on (126), it first occurred to us:

$$\text{a)}\ \frac{1}{2} - \frac{f'S}{f} = 0, \quad \text{i. e.} \qquad \text{b)}\ f = S^{1/2}. \tag{128}$$

This choice leads to the mass formula

$$m = m_0 \sqrt{1 + \frac{e^2 e_0^2 \, \mathrm{e}^{-3\sigma}}{m_0^2 c^4 S_0^2}} \tag{129}$$

(m_0 constant rest mass and e electric charge of the particle) and to the equation of motion

$$\frac{\mathrm{D}(mu^i)}{\mathrm{D}\tau} = \frac{e}{c} B^i{}_j u^j + \frac{3e^2 e_0^2 \, \mathrm{e}^{-3\sigma}}{2mc^2 S_0^2} \sigma^{|i} \tag{130}$$

with a scalaristic force term proportional to e^2. This term induces serious physical trouble. Concerning numerical values, the inequality

$$\frac{e^2 e_0^2 \, \mathrm{e}^{-3\sigma}}{m_0^2 c^4 S_0^2} \gg 1 \tag{131}$$

holds, i.e. the mass formula can hardly be accepted. This trouble already occurred in our former theory [4] and blocked our further investigation for more than two decades.

(ii) We also took into consideration the choice $f = \dfrac{\beta_0}{S}$ which leads to

$$\text{a)} \quad \frac{\mathrm{D}(mu^i)}{\mathrm{D}\tau} = \frac{e}{c} B^i{}_j u^j + mc^2 \sigma^{|i} - \frac{m}{2} \frac{\mathrm{d}\sigma}{\mathrm{d}\tau} u^i, \tag{132}$$

$$\text{b)} \quad m = m_0 \, \mathrm{e}^{-3\sigma/2}, \qquad \text{c)} \quad \beta = 1$$

with trouble of another kind.

(iii) Finally we decided to approach as close as possible the traditional equation of motion. This standpoint brought us to the choice

$$\text{a)} \quad \frac{1}{2} - \frac{3}{2}\beta^2 - \frac{f'S}{f} = 0, \quad \text{i.e.} \qquad \text{b)} \quad f = \sqrt{\frac{S_0 S}{1-\beta^2}} = \frac{1}{S}\sqrt{\beta_0^2 + S_0 S^3}\,. \tag{133}$$

Then the equation of motion (127) takes the well known shape

$$m_0 \frac{\mathrm{D}u^i}{\mathrm{D}\tau} = \frac{e}{c} B^i{}_j u^j \tag{134}$$

if we identify the electric charge as follows

$$e = \frac{\mathrm{i}\beta_0 m_0 c^2}{e_0 S_0}. \tag{135}$$

After we have used (125), the variational principle (121) acquires the 5-dimensional form

$$\delta \int_{P_0}^{P_1} S^{1/2} \sqrt{1 - \frac{e^2 e_0^2 S_0}{m_0^2 c^4 S^3}} \, \mathrm{d}\overset{5}{s} = 0 \tag{136}$$

which by means of (110) goes over into the 4-dimensional form

$$\delta \int\limits_{P_0}^{P_1} \left(1 - \frac{e^2 e_0^2 S_0}{m_0^2 c^4 S^3}\right) \mathrm{d}\overset{4}{s} = 0. \tag{137}$$

This last equation shows that for a particle without electric charge the geodesic motion applies.

5.11. Electrically charged perfect fluid

We try to treat the model of a medium consisting of uniform test particles with electric charge but without interaction. Of course, this model is an extreme idealization but rather adequate to this unified field theory of gravitation, electromagnetism and scalarism.

Up to now the substrate projector of the non-geometrized matter $\theta^{\mu\nu}$ played the role of a fully abstract quantity. In this chapter we shall investigate an ansatz for it for the case of this rather simple model which corresponds to a superposition of the test particles treated above. In constructing the substrate projector we will be supported by the above findings on the motion of the test particles.

Let us start with the most general ansatz*)

$$\theta^{\mu\nu} = A u^\mu u^\nu + B s^\mu s^\nu + C g^{\mu\nu} + D(u^\mu s^\nu + u^\nu s^\mu) \tag{138}$$

(A, B, C, D free coefficients). If we demand that by 4-bein projection the 4-dimensional energy tensor of the substrate

$$\text{a)}\ \theta^{ij} = -\left(\mu + \frac{p}{c^2}\right) u^i u^j - p g^{ij}, \qquad \text{b)}\ \theta^i{}_j = \mu c^2 - 3p \tag{139}$$

(μ mass density, p pressure) has to result, we obtain for the coefficients A and C:

$$\text{a)}\ A = -\frac{S\left(\mu + \dfrac{p}{c^2}\right)}{S_0(1-\beta^2)}, \qquad \text{b)}\ C = -\frac{S}{S_0}\, p. \tag{140}$$

Furthermore, the electric current density (74) takes the form of a convective current density

$$j^m = \varrho_0 u^m \tag{141}$$

with the electric charge density

$$\varrho_0 = \frac{\mathrm{i}\beta c^2 S^{3/2}}{e_0 S_0^{1/2} \sqrt{1-\beta^2}} \left[\mu + \frac{p}{c^2} + \frac{\mathrm{i} S_0 (1-\beta^2)}{c\beta S}\, D\right]. \tag{142}$$

It seems reasonable to assume that the pressure p does not induce electric charge. Hence follows

$$D = \frac{\mathrm{i}\beta S p}{c S_0 (1-\beta^2)}, \tag{143}$$

*) See new results in the Erice proceedings [7].

whereas the charge density takes the form

$$\varrho_0 = \frac{i\beta c^2 S^{3/2}}{e_0 S_0^{1/2} \sqrt{1-\beta^2}}\, \mu . \tag{144}$$

The most difficult problem is now how to find arguments for determining the last coefficient B. After long reflecting on this question we were led by the following idea:

Since a perfect fluid of the above type is apart from the 4-velocity characterized by the three physical properties μ, p and ϱ_0, we have to expect that according to (105) the scalarism inducing invariant $\left(\theta - \frac{3}{2}\,\theta^m{}_m\right)$ can in a certain way only be proportional to these quantities μ, p and ϱ_0:

(i) The mass density μ has to be excluded, because for a static field we would be left with

$$\Delta\sigma \sim \mu . \tag{145}$$

Furthermore, from (88) there would result a scalaristic force density of the structure

$${}^{(S)}f \sim \mu \operatorname{grad} \sigma . \tag{146}$$

Since the Newtonian gravitational theory, which has to be contained in our theory as the first approximation, means the Newtonian field equation (Φ Newtonian gravitational potential)

$$\Delta\Phi = 4\pi\gamma_N \mu \tag{147}$$

and the Newtonian gravitational force density

$${}^{(N)}\boldsymbol{f} = -\mu \operatorname{grad} \Phi , \tag{148}$$

we arrive at a parallelism of Φ and σ. This would mean that the empirical gravitational potential is a linear combination of Φ and σ, i.e. the relation (95) between Einstein's and Newton's gravitational constants would not hold.

(ii) We also exclude the charge density ϱ_0, because this version would according to (88) lead to a scalaristic force density of the type

$${}^{(S)}f^j \sim \varrho_0^2 \sigma^{|j} . \tag{149}$$

Passing over to the equation of motion of a test particle, we would arrive at a scalaristic force term, as already pointed out in (130). The physical trouble involved has already been discussed.

(iii) The only choice left and not being in contradiction to present experience seems to be that we admit the pressure p to induce scalarism and to appear in the scalaristic force density. This version leads to

$$B = \frac{\mu c^2 S}{2S_0(1-\beta^2)}\,(1 - 3\beta^2) + \frac{\beta^2 p S}{S_0(1-\beta^2)} . \tag{150}$$

Since our model of the fluid considered consists of the superposition of independent test particles, there must exist a fully logical consistency between the fluid theory and the test particle theory. Indeed, since the relations

$$\text{a) } \varrho_0 = en, \qquad \text{b) } \mu = m_0 n \quad (n \text{ particle density}) \tag{151}$$

are valid, with the help of (125), (133), (135) and (144) we find the coincidence

$$\frac{\varrho_0}{\mu} = \frac{e}{m_0} = \frac{i\beta_0 c^2}{e_0 S_0}. \tag{152}$$

The choice (150) determines the substrate projector as follows:

$$\theta^{\mu\nu} = \frac{e^\sigma}{1-\beta^2}\left[-\left(\mu + \frac{p}{c^2}\right) u^\mu u^\nu + \left\{\frac{\mu c^2}{2}(1-3\beta^2) + \beta^2 p\right\} s^\mu s^\nu \right.$$
$$\left. + \frac{i\beta p}{c}(u^\mu s^\nu + u^\nu s^\mu)\right] - e^\sigma p g^{\mu\nu}. \tag{153}$$

Hence results

$$\theta - \frac{3}{2}\theta^j{}_j = \frac{1}{2} p. \tag{154}$$

For this model the scalaristic field equation (105) takes the form

$$\sigma^{|k}{}_{;k} = \frac{1}{3}\varkappa_0 p + \frac{\varkappa_0}{8\pi} B_{kj} H^{kj}, \tag{155}$$

whereas the equation of motion (85) and the corresponding balance equation are:

$$\left(\mu + \frac{p}{c^2}\right)\frac{Du^m}{D\tau} + \left[\left(\mu + \frac{p}{c^2}\right)u^n\right]_{;n} u^m = -p^{|m} + \frac{1}{c} B^m{}_n j^n - \frac{1}{2} p\sigma^{|m}, \tag{156}$$

$$\left[\left(\mu + \frac{p}{c^2}\right)u^n\right]_{;n} = \frac{1}{c^2}\frac{dp}{d\tau} + \frac{1}{2c^2} p \frac{d\sigma}{d\tau}. \tag{157}$$

5.12. Variational principle

5.12.1. 5-dimensional variational principle

Let us introduce the field theoretical action in the 5-dimensional projective space and in the 4-dimensional space-time by [4]:

$$W = \frac{1}{c}\int_{V_5} \overset{5}{L}\, d^{(5)}f = \frac{1}{c}\int_{V_4} \overset{4}{L}\, d^{(4)}f \tag{158}$$

($\overset{5}{L}$ 5-dimensional Lagrange density, $\overset{4}{L}$ 4-dimensional Lagrange density). The 5-dimensional Hamilton principle reads

$$\text{a)}\ \delta \int_{V_5} \overset{5}{L}\, \mathrm{d}^{(5)}f = 0, \quad \text{where} \qquad \text{b)}\ \delta g_{\mu\nu}|_{(V_5)} = 0. \tag{159}$$

The variation of the 5-dimensional metric tensor yields the Lagrange equation

$$\frac{\delta \overset{5}{\mathscr{L}}}{\delta g_{\mu\nu}} = \frac{\partial \overset{5}{\mathscr{L}}}{\partial g_{\mu\nu}} - \left(\frac{\partial \overset{5}{\mathscr{L}}}{\partial g_{\mu\nu|\alpha}}\right)_{|\alpha} = 0 \tag{160}$$

if we restrict ourselves to Langrangians $\overset{5}{\mathscr{L}} = \overset{5}{L} \sqrt{\overset{5}{g}}$ of first order. According to

$$\overset{5}{\mathscr{L}} = \overset{(G)5}{\mathscr{L}} + \overset{(c)5}{\mathscr{L}} + \overset{(C)5}{\mathscr{L}} \tag{161}$$

the Lagrangian consists of the pure geometric term $\overset{(G)5}{\mathscr{L}}$, the cosmological term $\overset{(C)5}{\mathscr{L}}$ and the substrate term $\overset{(\theta)5}{\mathscr{L}}$, where

$$\text{a)}\ \overset{(G)5}{\mathscr{L}} = -\frac{\overset{5}{R}}{2\varkappa_0 S_0} \sqrt{\overset{5}{g}} + \text{divergence term}, \qquad \text{b)}\ \overset{(C)5}{\mathscr{L}} = \frac{\lambda_0 S}{\varkappa_0 S_0} \sqrt{\overset{5}{g}}\,. \tag{162}$$

Further, the relation

$$\theta^{\mu\nu} = -\frac{2S_0}{\sqrt{\overset{5}{g}}} \frac{\delta \overset{(\theta)5}{\mathscr{L}}}{\delta g_{\mu\nu}} \tag{163}$$

is valid.

If we take these facts into account, we find that the Lagrange equation (160) is identical with our basic field equation (50).

5.12.2. 4-dimensional variational principle

We are not able to prove here the relation

$$\overset{4}{L} = \frac{S_0^2}{S} \overset{5}{L}. \tag{164}$$

Using (162) we arrive at the 4-dimensional Langrangian

$$\overset{4}{\mathscr{L}} = \overset{4}{L} \sqrt{\overset{4}{g}} = \overset{(G)4}{\mathscr{L}} + \overset{(C)4}{\mathscr{L}} + \overset{(\theta)4}{\mathscr{L}} = \left(\overset{(G)4}{L} + \overset{(C)4}{L} + \overset{(\theta)4}{L}\right) \sqrt{\overset{4}{g}}\,, \tag{165}$$

where

$$\text{a)}\ \overset{(G)4}{L} = \frac{1}{2\varkappa_0} \overset{(Q)4}{R} - \frac{3}{4\varkappa_0 S^2} S_{|n} S^{|n} - \frac{1}{16\pi} B_{nk} H^{nk},$$

$$\text{b)}\ \overset{(C)4}{L} = \frac{\lambda_C}{\varkappa_0}, \qquad \text{c)}\ \overset{(\theta)4}{L} = \overset{(\theta f)4}{L} + \frac{1}{c} A_n j^n. \tag{166}$$

In these formulas $\overset{(Q)4}{R}$ means the quadratic part of the 4-dimensional curvature invariant, $\overset{(0f)4}{L}$ is the interaction free substrate part and A_n is the electromagnetic potential according to

$$B_{mn} = A_{n,m} - A_{m,n}. \tag{167}$$

Now the well-known 4-dimensional variational procedure is applied with the following result:

g_{ij} → generalized gravitational field equation (104),
A_n → inhomogeneous Maxwell equation (75),
S → scalaristic field equation (105).

Here the definition

$$\theta^{ij} = -\frac{2}{\sqrt{\overset{4}{g}}}\frac{\delta \overset{(\theta)4}{\mathscr{L}}}{\delta g_{ij}} \tag{168}$$

has to be used.

Thus the full consistency of the 5-dimensional and of the 4-dimensional variational principles is guaranteed.

Ladies and Gentlemen,

Please, let me summarize: I think we can optimistically look at our highly interesting field of physics which presents us with fresh challenges day by day. We are working at the roots of nature. We should gain new courage and strength when we reflect on Goethe's word in "Faust" written on this same historic soil, where our conference is taking place:

"that I might discover
the secret law
which holds the world
together at its core".

Literature

[1] Mercier, A.: In: „Albert Einstein — Sein Einfluß auf Physik, Philosophie und Politik" Edited by P. C. Aichelburg and R. U. Sexl. — Braunschweig—Wiesbaden: Friedr. Vieweg & Sohn GmbH. 1979.

[2] Schmutzer, E.: Nova Acta Leopoldina Halle (Suppl. No. 8) **44** (1976) 79; Exper. Technik Physik **24** (1976) 131; Acta Phys. Polon. **8B** (1977) 49; in: "Topics in Theor. and Exper. Gravitation Physics" (Erice Lectures 1975. Ed. V. De Sabbata and J. Weber) — New York: Plenum Press 1977; Gen. Rel. Grav. 8 (1977) 699 (Papapetrou-Festschrift).

[3] Koch, M.: Die Sterne **55** (1979) 52.

[4] Schmutzer, E.: Z. Physik **149** (1957) 329; **154** (1959) 312; Ann. Physik **1** (Folge 7) (1958) 136; Z. Naturforsch. **14a** (1959) 486, 489; **15a** (1960) 831; Forsch. u. Fortschr. **36** (1962) 321; Wiss. Z. Univ. Jena 8 (1958/59) 15; Relativistische Physik. — Leipzig: B. G. Teubner Verlagsgesellschaft 1968.
Kühnel, A.; Schmutzer, E.: Ann. Physik **7** (1961) 243.
[5] Jordan, P.: Schwerkraft und Weltall. — Braunschweig: Friedr. Vieweg & Sohn GmbH. 1955.
Ludwig, G.: Fortschritte der Projektiven Relativitätstheorie. — Braunschweig: Friedr. Vieweg & Sohn GmbH. 1951.
[6] Schmutzer, E.: Abstracts of GR9, Jena 1980; Exper. Technik Physik **28** (1980) 395; **28** (1980) 499; **29** (1981) 129; **29** (1981) 337; **29** (1981) 463.
[7] Schmutzer, E.: Exper. Technik Physik **30** (1982) 299; in: "Unified field theories of more than 4 dimensions" (Erice Lectures 1982. Ed. E. Schmutzer and V. De Sabbata) — Singapore: World Scientific Publ. Comp. 1983.
[8] Lichnérowicz, A.: Théories relativistes de la gravitation et de l'électromagnétisme. — Paris: Masson et C^ie 1955.

PLENARY LECTURES

Exact Solutions of Einstein's Field Equations

D. Kramer and H. Stephani (Jena)

Introduction

When preparing a short review on exact solutions, a field in which about 100 papers a year are published, one needs some selection rules. We decided firstly to concentrate on papers which appeared since 1974, with emphasis on the most recent years (1974 is the year of Kinnersley's report [0] on exact solutions, given at GR 7). Secondly, we agreed to give preference to the inhomogeneous cosmologies and to the stationary axisymmetric solutions because, in our opinion, the most interesting recent discoveries were made in these two fields.

Of course, this choice is also a matter of taste and inevitably reflects our own research interests.

Concerning the remaining classes of exact solutions, we can make only a few remarks. A more detailed presentation of what is known about exact solutions can be found in [1].

Starting point of all considerations are of course Einstein's famous field equations

$$R_{ab} - \frac{1}{2} R g_{ab} = \varkappa T_{ab}, \tag{1}$$

and in this review we shall deal only with the energy-momentum tensors for dust, perfect fluids, pure radiation, and electromagnetic fields (given in Table 1).

Table 1. *Energy-momentum tensors*

dust:	$T_{ab} = \mu u_a u_b, \qquad u_a u^a = -1,$
perfect fluid:	$T_{ab} = (\mu + p)\, u_a u_b + p g_{ab}$
pure radiation:	$T_{ab} = \Phi^2 k_a k_b, \qquad k_a k^a = 0,$
electromagnetic field:	$T_{ab} = F_{ac} F_b{}^c - \frac{1}{4} g_{ab} F_{cd} F^{cd}.$

1. Homogeneous and inhomogeneous cosmologies

Einstein's theory has an important impact on cosmology. The cosmic matter content is usually assumed to be dust, or a perfect fluid with an equation of state

$$p = (\gamma - 1)\, \mu, \qquad \gamma = \text{const}, \tag{2}$$

and the cases most frequently considered are incoherent radiation ($p = \mu/3$) and stiff matter ($p = \mu$). A perfect fluid, or dust solution of Einstein's field equations will be called a cosmological model or, simply, a cosmology.

In the well-known Robertson-Walker-Friedmann (RWF-) models, all space points and all directions at any space point are equivalent. These spatially-homogeneous and isotropic models fit the experimental data very well. The isotropy on a large scale has been confirmed by observations of the microwave background. But was the universe homogeneous and isotropic already at its early stages, and has it still these properties in very distant regions? Some phenomena suggest that anisotropic models could provide a more appropriate description of the early stages than the RWF-models. Physical processes such as particle creation might have damped anisotropy. There are also exact anisotropic and inhomogeneous perfect fluid solutions which evolve towards RWF-models [2].

A cosmology is either *homogeneous*, or *spatially-homogeneous*, or (spatially) *inhomogeneous*. A homogeneous cosmology admits a group of motions acting on space-time; the group orbits of a spatially-homogeneous cosmology are spacelike hypersurfaces. All other cosmologies are said to be inhomogeneous.

1.1. Homogeneous cosmologies

All perfect fluid solutions, homogeneous in space and time, are known [3, 4], but they can hardly provide a realistic picture of the universe. As recently shown [5], the only homogeneous vacuum solutions are special plane waves, and the Petrov solution [6]

$$k^2\,\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{e}^{-2x}\,\mathrm{d}y^2 + \mathrm{e}^{x}\left[\cos\sqrt{3}\,x(\mathrm{d}z^2 - \mathrm{d}t^2) - 2\sin\sqrt{3}\,x\,\mathrm{d}z\,\mathrm{d}t\right] \tag{3}$$

(k constant). The solution (3) admits a simply-transitive group G_4 and has been interpreted as the field in the interaction region of colliding plane waves [7], or as the exterior gravitational field of a cylindrically symmetric, stationarily rotating dust source [8].

The only homogeneous Einstein-Maxwell fields where the electromagnetic field shares the space-time symmetry are special plane waves, and the Bertotti-Robinson solution [9, 10]

$$k^2\,\mathrm{d}s^2 = \mathrm{d}\vartheta^2 + \sin^2\vartheta\,\mathrm{d}\varphi^2 + \mathrm{d}x^2 - \sinh^2 x\,\mathrm{d}t^2 \tag{4}$$

(k constant), which plays a role in studies of colliding plane waves in the Einstein-Maxwell theory [11].

The existence of a space-time symmetry implies

$$\mathfrak{L}_\xi F_{ab} = C\tilde{F}_{ab} \tag{5}$$

for the Lie derivative of the Maxwell tensor F_{ab} with respect to the Killing vector ξ^a. ($\tilde{F}_{ab}$ denotes the dual Maxwell tensor.) C is a constant for non-null electromagnetic fields [14] and the gradient $C_{,a}$ must be proportional to the repeated principal null vector of a null field [15].

There are homogeneous Einstein-Maxwell fields where the electromagnetic field has lower symmetry than the geometry. An example is [12, 13]

$$ds^2 = \frac{a^2}{x^2}(dx^2 + dy^2) + x^2\,d\varphi^2 - (dt - 2y\,d\varphi)^2, \quad a \text{ constant}, \tag{6}$$

$$\sqrt{\varkappa}\,F_{41} = \frac{2}{x}\cos 2\ln x, \quad \sqrt{\varkappa}\,F_{13} = \frac{4y}{x}\cos 2\ln x, \quad \sqrt{\varkappa}\,F_{23} = 2\sin 2\ln x,$$

with $x^a = (x, y, \varphi, t)$. For the Killing vector $\xi = x\,\partial_x + y\,\partial_y - \varphi\,\partial\varphi$ one obtains $C = 2$ in (5). By the way, (6) is a Petrov type I solution and the principal null directions of the Weyl and Maxwell tensors are *not aligned.*

1.2. Spatially-homogeneous cosmologies

Next we come to the anisotropic spatially-homogeneous cosmologies. The group of motions is either a (multiply-transitive) four-parameter group or a (simply-transitive) three-parameter group. Since in the former case there is a one-parameter group of local rotations, these spaces are called *locally rotationally symmetric* (L.R.S.). This rotational symmetry is a remnant of the three-parameter isotropy of the RWF-models.

Atll but one of the L.R.S. metrics contain a simply-transitive subgroup. The exceptional case

$$ds^2 = -dt^2 + A^2(t)\,dx^2 + B^2(t)\,(dy^2 + \sin^2 y\,dz^2) \tag{7}$$

has been investigated by Kantowski and Sachs [16], and other authors.

The cosmologies admitting a simply-transitive group are the so-called Bianchi-type models. They can jointly be described by the metric

$$ds^2 = -dt^2 + g_{\alpha\beta}(t)\,\omega^\alpha\omega^\beta, \quad \alpha, \beta = 1, \ldots, 3, \tag{8}$$

where ω^α are three (time-independent) basis one-forms invariant under the group. The isotropic models are contained as special cases, e.g.

$$\omega^\alpha = dx^\alpha, \quad g_{\alpha\beta} = a^2(t)\,\delta_{\alpha\beta}. \tag{9}$$

The metrics (7) and (8) cover all spatially-homogeneous models. Cosmologies in which the fluid 4-velocity is not orthogonal to the homogeneous space slices, the group orbits, are said to be *tilted.*

The field equations lead to sets of *ordinary* differential equations. Nevertheless, explicit analytical solutions are known only for some cases, especially for an equation of state $p = (\gamma - 1)\,\mu$, $\gamma = 1, 4/3, 2$, and much work remains to be done. For illustration we give as an example of an exact Bianchi-type solution the metric [17, 18]:

$$ds^2 = -dt^2 + \frac{t^2\,dx^2}{(m-n)^2} + \frac{1}{t^{2(m+n)}}(e^{-2x}\,dy^2 + e^{2x}\,dz^2),$$

$$\varkappa\mu = 4\,\frac{m^2 + mn + n^2}{t^2}, \quad \varkappa p = -\frac{4mn}{t^2}, \tag{10}$$

$$2m^2 + 2n^2 + m + n = 0, \quad m, n \text{ constants}.$$

Table 2. *Contributions to the spatially-homogeneous cosmologies (since 1970)*

$p = (\gamma - 1)\,\mu$	Collins [17], Dunn and Tupper [18], Vajk and Eltgroth [19]
Λ-term	Siklos [20], MacCallum and Siklos [21]
dust	Evans [22]
stiff matter ($p = \mu$)	Send [23], Maartens and Nel [24], Barrow [25], Ruban [26], Wainwright et al. [27]

Some papers from the last decade are listed in Table 2. In particular, Wainwright et al. [27] derived a variety of both new tilted spatially-homogeneous and new inhomogeneous cosmologies with irrotational stiff matter, see also section 1.3.4. One of these exact solutions has been interpreted as a gravitational wave pulse of finite duration moving through a Bianchi type model.

If the matter content is not a perfect fluid but an electromagnetic field, new solutions are given by Barnes [28] who generalized an ansatz due to Tariq and Tupper [29]. The essential assumption is that the derivatives of the (complex self-dual) Maxwell tensor along its null eigendirections are proportional to this tensor. It is surprising that this assumption implies the existence of a three-parameter group of motions. The Barnes solutions contain new spatially-homogeneous vacuum models as special cases.

Some solutions with perfect fluid *and* electromagnetic field have been studied, e.g. [30—32]. For large cosmological time, the model considered by Damião Soares [32] asymptotically approaches a perfect fluid solution with the equation of state $p = \mu/5$.

As far as we are aware, none of the known anisotropic statially-homogeneous cosmologies is favoured by observational evidence.

If the group of motions acts on *timelike*, not on spacelike hypersurfaces, then corresponding similar solutions exist. For an Abelian group, they usually are interpreted as stationary cylindrically symmetric solutions.

In the case of three-dimensional *null* orbits it has been shown [33] that if the energy conditions are satisfied, there is a non-expanding and shearfree geodesic null congruence which is a common eigendirection of both the Weyl and Ricci tensors. The corresponding space-times are algebraically special.

1.3. Inhomogeneous cosmologies

The inhomogeneous cosmologies include two physically important types. The first type could be termed stellar models; they possess a closed 2-surface of vanishing pressure and could in principle be matched to an exterior vacuum solution. The second type justifies the name cosmology in so far as these solutions resemble the real universe in some aspects and can at least be considered as a reasonable description of parts of the universe.

1.3.1. Spherically symmetric solutions

All spherically symmetric *dust* solutions have been known for a long time; this class is due to Tolman [34]. Large classes of *static* perfect fluid solutions have been found, and numerical methods are available. A disadvantage of most of the analytical methods for finding solutions is that the equation of state cannot be prescribed at the beginning, but has to be calculated from the resulting metric.

The treatment of the *non-static* case depends very much on the shear of the fluid.

If the radial motion is *non-shearing*, i.e. if a moving volume element does not change its shape, then the field equations reduce to the single equation [35]

$$L_{,xx} = L^2(x, t)\, F(x), \quad L_{,t} \neq 0. \tag{11}$$

Once a solution L is found, the metric can be calculated via

$$\begin{aligned} ds^2 &= e^{2\lambda(r,t)}(r^2\, d\vartheta^2 + r^2 \sin^2 \vartheta\, d\varphi^2 + dr^2) - \dot{\lambda}^2\, e^{-2f(t)}\, dt^2, \\ L &\equiv e^{-\lambda}, \qquad x \equiv r^2, \end{aligned} \tag{12}$$

$F(x)$ and $f(t)$ being arbitrary functions. All known non-static shearfree spherically symmetric perfect fluids belong to the class

$$F(x) = (ax^2 + bx + c)^{-5/2} \tag{13}$$

discovered by Kustaanheimo and Qvist [35].

In the case of a *shearing* radial motion, only a few rather special solutions have been found so far, see [36—39]. So, perhaps contrary to what the outsider might expect, we have to state that only a minority of the spherically symmetric perfect fluid solutions are actually known.

1.3.2. Plane symmetric solutions

All plane symmetric *dust* solutions are known; they are contained in the general spherically symmetric dust solutions.

Plane symmetric *perfect fluid* solutions have been found less attention than the corresponding solutions with spherical symmetry. So even in the static case, only a few solutions have been found. Cases which have been treated are, e.g., $\mu =$ const [40, 41]; $\mu = p$ [42, 27]; $\mu = 3p$ [43]. For illustration, we give the static solution with $\mu = 3p$; it reads

$$\begin{aligned} ds^2 &= z^2(dx^2 + dy^2) + zF^{-1}\, dz^2 - p^{-1/2}\, dt^2, \\ 5F &= \varkappa p_0(216 - 108z^5 + 18z^{10} - z^{15}), \\ p &= p_0(36z^2 - 12z^7 + z^{12}). \end{aligned} \tag{14}$$

1.3.3. Stationary axisymmetric solutions

The stationary axisymmetric perfect fluid solutions should provide the most simple models of the interior of a rotating star. Such models require a physically acceptable equation of state, and a matching to an exterior vacuum solution. None of the known solutions seem to meet these two requirements. The interior (perfect fluid) Kerr solution is still an unsolved problem! But let us have a look at the solved problems.

All known perfect fluid and dust solutions obey the circularity condition, which ensures that the 4-velocity is coplanar with the Killing vectors.

All stationary rotating *dust* metrics of this type are known up to quadratures. This result, which covers the general case of differential rotation, was obtained by Winicour in 1975 [44]. In order to construct these metrics, one has to choose a real function of one variable, and an axisymmetric solution of the potential equation in flat 3-space. Then the metric functions, the angular velocity, and the mass density can be determined by means of line integrals and elementary operations. In principle, all solutions of this class are known, but since the procedure of calculating the metric is rather involved, the physical properties of rotating dust are in fact only incompletely known. In Newtonian gravity, an *isolated*, axisymmetric, steadily rotating dust cloud cannot exist: any density gradient (or jump of density) in the direction of the rotation axis would cause a motion in that direction. In Einstein's theory, a similar result is to be expected.

A subclass of cylindrically symmetric Winicour solutions, including the rigidly rotating van Stockum cylinder [45], has the remarkable property that it can be matched to a *static* exterior vacuum solution [45, 46].

In the perfect fluid case, we know only one stationary axisymmetric solution without higher symmetry: the rigidly rotating fluid solution found by Wahlquist [47]. It obeys an equation of state $\mu + 3p = \text{const.}$

1.3.4. Inhomogeneous cosmologies with irrotational stiff matter

Progress has been made in constructing inhomogeneous cosmological models with *irrotational stiff* matter:

$$u_a = \frac{\sigma_{,a}}{\sqrt{-\sigma_{,c}\sigma^{,c}}}, \qquad p = \mu. \tag{15}$$

It is assumed that an Abelian group G_2 acts on spacelike orbits and that 2-surfaces orthogonal to these orbits exist, i.e., G_2 is an orthogonally transitive group. In this case, the metric can be put into the form

$$ds^2 = e^M(dz^2 - dt^2) + g_{AB}\, dx^A\, dx^B, \quad A, B = 1, 2, \tag{16}$$

where the metric coefficients depend only on x^A. The structure of the field equations

$$R_{ab} = 2\varkappa\sigma_{,a}\sigma_{,b}, \qquad p = \mu = -\sigma_{,c}\sigma^{,c}, \tag{17}$$

allows one to generate perfect fluid solutions of the type under consideration from known vacuum solutions, simply by replacing M in (16) by a new function which is to be calculated via a line integral [42, 27].

1.3.5. Inhomogeneous cosmologies with conformally flat slices

Now let us turn to another class of inhomogeneous cosmologies. In his 1975 paper [48], Szekeres very successfully started considering metrics of the form

$$\begin{aligned} \mathrm{d}s^2 &= \mathrm{e}^{2A}\,\mathrm{d}r^2 + \mathrm{e}^{2B}(\mathrm{d}x^2 + \mathrm{d}y^2) - \mathrm{d}t^2, \\ A &= A(x, y, r, t), \quad B = B(x, y, r, t), \quad u^a = (0, 0, 0, 1). \end{aligned} \tag{18}$$

It was discovered only later that these metrics can be characterized by the existence of conformally flat slices $t = \text{const}$. Szekeres found all dust metrics of this class, and, among others [49—52, 70], dealt with the perfect fluid case. Szafron and Wainwright [51] presented a new class of inhomogeneous and anisotropic cosmologies, which, in the limit of large cosmological time, approximate a RWF solution.

For the integration procedure of the field equations, two cases have to be distinguished.

If *B depends on r*, then the metric must have the form

$$\mathrm{d}s^2 = \Phi^2(r, t)\,[P^{-2}(\mathrm{d}x^2 + \mathrm{d}y^2) + (\partial_r \ln\{\Phi P^{-1}\})^2\,\mathrm{d}r^2] - \mathrm{d}t^2 \tag{19}$$

with x-y-spaces of constant curvature K:

$$\begin{aligned} P(x, y, r) &= a(r)\,(x^2 + y^2) + b(r)\,x + c(r)\,y + \mathrm{d}(r), \\ K &= 4ad - b^2 - c^2. \end{aligned} \tag{20}$$

The function Φ is a solution of the ordinary (Friedmann-type) differential equation

$$2\Phi\ddot{\Phi} + \dot{\Phi}^2 + \varkappa p(t)\,\Phi^2 = 1 - K(r). \tag{21}$$

To get an explicit solution, one has to prescribe the pressure p and the four real functions a, b, c, d and then to solve (21).

For dust ($p = 0$), this differential equation can be completely solved. The solutions contain as special cases the Friedmann and Tolman dust solutions. In these spherically symmetric cases, the radially moving dust clouds can be thought to form concentric shells. In the corresponding general Szekeres solution, the different dust shells would have different centres, and the metric admits no Killing vector. Surprisingly, these solutions can be matched to the exterior Schwarzschild solution! They thus provide us with an example of a non-spherically symmetric system of particles in free gravitational motion which does not emit gravitational radiation. This remarkable result is due to Bonnor [53].

If *B does not depend on r*, then the metric is

$$\begin{aligned} \mathrm{d}s^2 = \Phi^2(t)\big[&P^{-2}(x, y)\,(\mathrm{d}x^2 + \mathrm{d}y^2) \\ &+ \big\{C(r, t)\,P^{-1}\big(U(r)\,(x^2 + y^2) + V(r)\,x + W(r)\,y + Z(r)\big)\big\}\,\mathrm{d}r^2\big] - \mathrm{d}t^2 \end{aligned} \tag{22}$$

with

$$\begin{aligned} &P = 1 + k(x^2 + y^2)/4, \qquad k = 0, \pm 1, \\ &2\Phi\ddot{\Phi} + \dot{\Phi}^2 + \varkappa p(t)\,\Phi^2 = -k, \\ &\ddot{C}\Phi^2 + 3\dot{C}\dot{\Phi}\Phi - Ck = 2U + kZ. \end{aligned} \tag{23}$$

To get an explicit solution, one has to specify the pressure $p(t)$ and the functions $U(r)$, $V(r)$, $W(r)$ and $Z(r)$ and then to determine $\Phi(t)$ and $C(r, t)$ from (23).

For dust, the general solution was given by Szekeres. It contains the Kantowski-Sachs models as special cases. The evolution of these solutions has been studied by Bonnor and Tomimura [54] with the result that some of them become spatially-homogeneous for t approaching infinity.

1.3.6. Algebraically special inhomogeneous cosmologies

We close this chapter with a few remarks on the Petrov types of the known inhomogeneous cosmologies. Most of the solutions considered above are of type D, e.g. the spherically (or plane) symmetric solutions, the Wahlquist solution, and the Szekeres class with conformally flat slices. Making explicit use of the existence of a geodesic and shearfree (or twistfree) multiple null eigenvector, Wainwright [76, 77] and Oleson [78, 79] constructed several classes of algebraically special perfect fluid solutions. In general, they do not have an obvious physical interpretation; their Petrov types are II, D, N, and O. Apparently, no type III perfect fluid (or dust) solution is known. In the type N solutions, the fluid must have non-zero acceleration [79].

2. Stationary axisymmetric vacuum and Einstein-Maxwell fields

It is very attractive to study the stationary axisymmetric vacuum and electrovacuum fields because they include the exterior gravitational fields of rotating isolated mass and charge distributions. Moreover, the field equations can be reduced to a fascinating symmetric form when expressed in terms of two complex potentials: the Ernst and the electromagnetic potentials. In the last few years, successful and exciting studies were devoted to this problem. Now there are new powerful methods which enable one to construct reasonable solutions with any number of parameters, and it is very likely that even the general solution will be found on the basis of these recent developments.

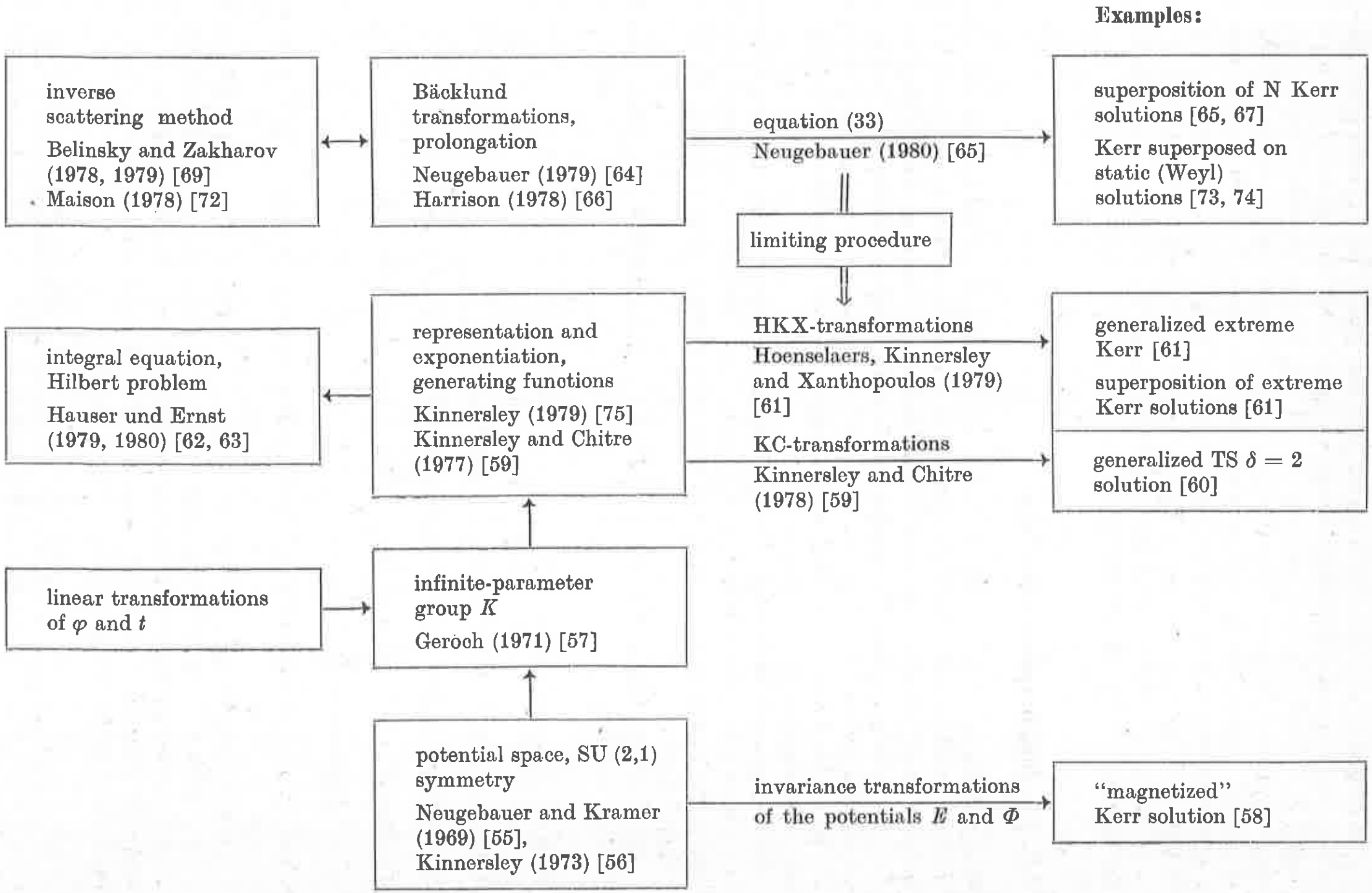
In the following sections we shall give a survey of these methods and the main results. Table 3 shows how (and when) these generation methods have developed and how they are interrelated.

All the results mentioned in this section also apply, with slight modifications, to the (non-stationary) *cylindrically* symmetric fields.

2.1. The SU(2,1) internal symmetry

Prior to 1974 the Tomimatsu-Sato class covered all known vacuum fields, including the Kerr solution, which might be produced by isolated rotating massive sources. Furthermore, one knew of a solution-generating procedure for stationary axisymmetric Einstein-Maxwell fields outside the sources. This method can be summarized as follows. The Ernst and electromagnetic potentials, and their complex conjugates, can

Table 3. *Generation methods for stationary axisymmetric vacuum and Einstein-Maxwell fields*



be considered as local coordinates in the potential space. Due to an internal symmetry of the field equations, this potential space admits an SU(2,1) symmetry group [55, 56] which, for vacuum fields, contains an SU(1,1) subgroup [57]. The corresponding finite invariance transformations preserve the potential space metric and the form of the field equations. By means of such transformations, which contain at most 8 real parameters, new solutions were obtained from previously known ones. For instance, electrified versions of the Tomimatsu-Sato solutions were generated. Another example is an exact solution for a Kerr black hole embedded in an asymptotically homogeneous magnetic field [58].

2.2. The infinite-parameter group

The starting point for further investigations was the idea that one can successively perform two kinds of invariance operations: the internal symmetry already referred to, and linear transformations of the ignorable coordinates φ and t associated with the two Killing vectors. The product of these two non-commuting Lie groups forms an infinite-parameter group, K. In order to represent the infinitesimal actions of K, Kinnersley and Chitre [59] introduced an infinite hierarchy of potentials, one of them being the Ernst potential.

For some special subgroups of K which do preserve asymptotic flatness, Kinnersley and Chitre [59, 60] succeeded in exponentiating the infinitesimal transformations. The finite transformations of one of these subgroups yield generalized Tomimatsu-Sato-solutions.

For instance, the Ernst potential E of the generalized $\delta = 2$ Tomimatso-Sato solution is given as

$$
\begin{aligned}
E &= \frac{1-\xi}{1+\xi}, \qquad \xi = \frac{\alpha}{\beta}, \\
\alpha &= p^2(x^4-1) - 2ipqxy(x^2-y^2) + q^2(y^4-1) \\
&\quad -2ia(x^2+y^2-2x^2y^2) - 2ibxy(x^2+y^2-2) + (a^2-b^2)(x^2-y^2)^2, \\
\beta &= 2px(x^2-1) - 2iqy(1-y^2) - 2i(pa+iqb)\,x(x^2-y^2) \\
&\quad -2i(pb+iqa)\,y(x^2-y^2),
\end{aligned}
\tag{24}
$$

where (x, y) are prolate spheroidal coordinates and p, q $(p^2+q^2=1)$, a, b are parameters.

Another subgroup gives rise to the *HKX-transformations* [61]. They can be applied to any given solution. When applied to Minkowski space, the simplest members of these transformations yield an asymptotically flat generalization of *extreme* Kerr. For a special choice of parameters, the corresponding gravitational potential ξ is given, in terms of spherical coordinates (r, ϑ), in the form

$$
\xi = \frac{c_1r^3 - 2c_2r^2\cos\vartheta - 2ic_2{}^2\cos\vartheta\sin^2\vartheta}{r^4 - ic_1r^3\cos\vartheta + 2ic_2r^2(\cos^2\vartheta-\sin^2\vartheta) - c_2{}^2\sin^2\vartheta(1+\cos^2\vartheta)} \tag{25}
$$

(c_1, c_2 constants). When higher-rank transformations are included, the tedious, but straightforward calculations lead to Ernst potentials of rapidly increasing complexity.

The common feature of the two kinds of finite transformations which have been found so far by exponentiation is that they preserve asymptotic flatness.

To master the infinite hierarchy of potentials (which represents a very redundant description) and to find the exponentiations, it proved to be convenient, if not necessary, to work with generating functions from which all the potentials can be derived. These generating functions depend on at least one additional variable besides the two non-ignorable space-time coordinates.

The Kinnersley-Chitre representation of K allows an interesting reformulation due to Hauser and Ernst [62]: the corresponding finite transformations can be effected by solving a linear integral equation of the Cauchy type. This integral equation in turn is equivalent to a homogeneous *Hilbert problem* which is dealt with in complex function theory. In the vacuum case, this problem can be stated as follows [63].

Let L be a smooth contour surrounding the origin in a complex s-plane. Find two 2×2 matrix functions $X_+(s)$ and $X_-(s)$ such that

(i) $X_+(s)$ is holomorphic on and within L,

(ii) $X_-(s)$ is holomorphic on and outside L,

(iii) $X_+(0) = \mathrm{I} =$ unit 2×2 matrix,

(iv) the inverses of $X_+(x)$ and $X_-(s)$ exist,

(v) the *boundary values on L* obey

$$X_-(s) = X_+(s)\, G(s), \quad G(s) = F(s)\, u(s)\, F^{-1}(s), \tag{26}$$

where the 2×2 complex matrix functions $u(s)$ satisfy the relations

$$\det u(s) = 1, \qquad u(s)^+\, \varepsilon\, u(s) = \varepsilon,$$
$$\varepsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad u(s)^+ = \text{hermitean conjugate of } u(\bar{s}), \tag{27}$$

$F(s)$ is a 2×2 matrix generating function associated with a given vacuum solution,

(vi) certain conditions at infinity are imposed.

Once this Hilbert problem has been solved, for given $F(s)$ and $u(s)$, one obtains a new vacuum solution from the generating matrix function $F'(s) = X_+(s)\, F(s)$. The general solution of the Hilbert problem in several unknown functions is not known in closed form. Nevertheless, this new formulation might be helpful in gaining deeper insights into the underlying mathematical structure of the field equations.

2.3. The method of Bäcklund transformations

Now we turn to another very successful approach: the extension of methods used in soliton physics to Einstein's theory of stationary axisymmetric fields. Again we will restrict ourselves to the vacuum case.

The space-time metric can be written in the standard form

$$ds^2 = e^{-2U}(e^{2k}\,dz\,d\bar{z} + W^2\,d\varphi^2) - e^{2U}(dt + \omega\,d\varphi)^2, \tag{28}$$

where all functions depend on the complex conjugate coordinates z and $\bar{z}$. The relevant field equations are (in three-dimensional vector notation)

$$W_{,z\bar{z}} = 0, \quad (\mathrm{Re}\,E)\,\Delta E = (\nabla E)^2, \quad E \text{ Ernst potential.} \tag{29}$$

Introducing new variables

$$\begin{aligned} A_1 &= \frac{E_{,z}}{E+\bar{E}}, \quad & B_1 &= \frac{\bar{E}_{,z}}{E+\bar{E}}, \quad & C_1 &= \frac{W_{,z}}{W}, \\ A_2 &= \frac{E_{,\bar{z}}}{E+\bar{E}}, \quad & B_2 &= \frac{\bar{E}_{,\bar{z}}}{E+\bar{E}}, \quad & C_2 &= \frac{W_{,\bar{z}}}{W}, \end{aligned} \tag{30}$$

the field equations (29) can be cast into the system

$$\begin{aligned} A_{1,\bar{z}} &= -A_1B_2 + A_1A_2 - \frac{1}{2}\,C_1A_2 - \frac{1}{2}\,C_2A_1 \\ A_{2,z} &= -A_2B_1 + A_1A_2 - \frac{1}{2}\,C_1A_2 - \frac{1}{2}\,C_2A_1 \\ B_{1,\bar{z}} &= -B_1A_2 + B_1B_2 - \frac{1}{2}\,C_1B_2 - \frac{1}{2}\,C_2B_1 \\ B_{2,z} &= -B_2A_1 + B_1B_2 - \frac{1}{2}\,C_1B_2 - \frac{1}{2}\,C_2B_1 \\ C_{1,\bar{z}} &= -C_1C_2 \\ C_{2,z} &= -C_1C_2 \end{aligned} \tag{31}$$

of first-order differential equations.

For $W = 1$, these equations reduce to a subsystem connected with the sine-Gordon equation

$$u_{,z\bar{z}} = \sin u \tag{32}$$

(in two space dimensions) which plays an important role in various areas of physics, e.g. superconductivity theory. With the aid of Bäcklund transformations (BT) one can generate new solutions of this equation, by purely algebraic manipulations.

A look at the differential equations (31) suggests that the BT of the sine-Gordon equation, i.e. of the subsystem with $C_1 = C_2 = 0$, should have a generalized version for Einstein's theory of stationary axisymmetric vacuum field, i.e. for the total system. This gravitational counterpart does indeed exist [64—66]. The repeated application of these gravitational Bäcklund transformations to a known solution (e.g. flat space) leads to new solutions with any number of parameters.

Let us have a look at the final formula for the Ernst potential E' of a new solution which is obtained, after an even number of recursion steps, from an initial solution with the Ernst potential E

$$E' = E \frac{\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ a_0 & a_1\zeta_1 & a_2\zeta_2 & \cdots & a_n\zeta_n \\ 1 & \zeta_1^2 & \zeta_2^2 & \cdots & \zeta_n^2 \\ a_0 & a_1\zeta_1^3 & a_2\zeta_2^3 & \cdots & a_n\zeta_n^3 \\ \vdots & \vdots & \vdots & & \vdots \\ a_0 & a_1\zeta_1^{n-1} & a_2\zeta_2^{n-1} & \cdots & a_n\zeta_n^{n-1} \\ 1 & \zeta_1^n & \zeta_2^n & \cdots & \zeta_n^n \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & a_1\zeta_1 & a_2\zeta_2 & \cdots & a_n\zeta_n \\ 1 & \zeta_1^2 & \zeta_1^2 & \cdots & \zeta_n^2 \\ 1 & a_1\zeta_1^3 & a_2\zeta_2^3 & \cdots & a_n\zeta_n^3 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & a_1\zeta_1^{n-1} & a_2\zeta_2^{n-1} & \cdots & a_n\zeta_n^{n-1} \\ 1 & \zeta_1^n & \zeta_2^n & \cdots & \zeta_n^n \end{vmatrix}} \tag{33}$$

(Neugebauer [65]). This closed-form expression for E' contains only E, the quantities ζ_k defined by

$$\zeta_k = \left(\frac{c_k - \mathrm{i}\bar{z}}{c_k + \mathrm{i}z}\right)^{1/2}, \quad c_k \text{ constants}, \tag{34}$$

and the solutions a_k of the total Riccati equation

$$\begin{aligned} a_{,z} &= \frac{1}{E + \bar{E}} \left[(a - \zeta)\, \bar{E}_{,z} + (a^2\zeta - a)\, E_{,z}\right], \\ a_{,\bar{z}} &= \frac{1}{E + \bar{E}} \left[\left(a - \frac{1}{\zeta}\right) \bar{E}_{,\bar{z}} + \left(\frac{a^2}{\zeta} - a\right) E_{,\bar{z}}\right] \end{aligned} \tag{35}$$

with different integration constants indicated by the index k. At each recursion step further integration constants appear. They can be chosen such that the reality of the metric and asymptotic flatness are preserved.

The total Riccati equation (35) can be linearized by the ansatz $a = -\psi/\chi$ which gives rise to the linear eigenvalue equations

$$\begin{aligned} \psi_{,z} &= B_1(\psi + \zeta\chi), \qquad \psi_{,\bar{z}} = B_2\left(\psi + \frac{1}{\zeta}\,\chi\right), \\ \chi_{,z} &= A_1(\chi + \zeta\psi), \qquad \chi_{,\bar{z}} = A_2\left(\chi + \frac{1}{\zeta}\,\psi\right), \end{aligned} \tag{36}$$

where the eigenvalue is hidden in ζ. The two complex potentials ψ and χ are closely related to the matrix generating function F mentioned above.

The formula (33) provides us with a simple recipe for constructing new solutions. To our knowledge, all stationary axisymmetric solutions given hitherto in the literature can be obtained from this formula (including limiting procedures).

The *application* of the formula (33) to flat space ($E = 1$) leads to a non-linear superposition of Kerr-NUT solutions [65, 67]. It is a tempting task to find out whether the gravitational attraction of corotating Kerr particles can be balanced by rotational repulsion.

For Weyl's static class, the Riccati equation for a can be solved exactly, and (33) yields Kerr- NUT superposed on static solutions [73, 74].

At the limit where two or more columns of the generalized Vandermonde determinants in the numerator and denominator of (33) coincide, one arrives at the HKX-transformations. For instance, the solution (25) can be rediscovered as such a limiting case.

It has been shown by Cosgrove [68] that the method of BT is equivalent to the approach developed by Belinsky and Zakharov [69]. These authors extended the *inverse scattering method* to Einstein's theory; they started from linear eigenvalue equations, with spectral parameter λ, and solved them by expansions in terms of poles in the complex λ-plane. The so-called soliton transformations thus obtained correspond to the BT and, at the limit of coalescing poles, to the HKX-transformations. Belinsky and Zakharov did not derive an explicit formula for calculating the new metric. The non-soliton part of the inverse scattering transform gives rise to a linear integral equation, and an associated Hilbert problem, which is similar but not equivalent to that stated above.

Some of the approaches listed in Table 3 have been generalized to include electromagnetic fields, see e.g. [62, 63, 71].

Herlt [97] has recently been able to generate a new class of asymptotically flat static Einstein-Maxwell fields in closed form. This class contains a disposable axisymmetric solution of the potential equation in flat 3-space and goes over to the Schwarzschild solution when the electric field is switched off.

3. Algebraically special solutions

Algebraically special solutions are characterized by the existence of (at least) one repeated null eigenvector k^a of the Weyl tensor. The techniques of constructing solutions depend essentially on the properties of the corresponding null vector field, especially on the complex divergence $\varrho = -(\theta + i\omega)$. For Einstein-Maxwell fields it is also important whether or not the eigenvector k^a is parallel, or *aligned*, to an eigenvector of the Maxwell tensor.

Table 4 shows the present status of the algebraically special vacuum, Einstein-Maxwell, and pure radiation fields. It also indicates for which classes new results have been obtained since 1974. The symbols mean: $\nexists$: solutions do not exist, A: All solutions are known, S: Some solutions are known.

Table 4. *Present status of the algebraically special solutions, and the relevant papers since 1974. Solutions may not exist* ($\nexists$), *or some solutions* (S) *or all solutions* (A) *a known. See text for further explanation.*

Energy-momentum-tensor ↓	Petrov type →	II	D	III	N	O
Vacuum	$\varrho \neq \bar{\varrho}$	S Robinson [80]	A	S Held [81] Robinson [80]	S Hauser [82]	$\nexists$
	$\varrho = \bar{\varrho} \neq 0$	S	A	S	A	$\nexists$
	$\varrho = 0$	S	A	A	A	$\nexists$
Einstein-Maxwell nonnull (aligned, $\varkappa = 0$)	$\varrho \neq \bar{\varrho}$	S	S Plebański-Demiański [83] Leroy [84]	$\nexists$	$\nexists$	$\nexists$
	$\varrho = \bar{\varrho} \neq 0$	S Leroy [85] Kowalczyński [98]	A Leroy [85]	$\nexists$	$\nexists$	$\nexists$
	$\varrho = 0$	S	S Kowalczyński-Plebański [86]	S Hacyan-Plebański [87]	A	A
(non-aligned)				S	A	$\nexists$
Einstein-Maxwell null	$\varrho \neq \bar{\varrho}$				$\nexists$	$\nexists$
	$\varrho = \bar{\varrho} \neq 0$	S	A	S	$\nexists$	$\nexists$
	$\varrho = 0$	S			S Debney [88]	A
pure radiation		S Herlt [89] Stephani [90]	S Frolov-Khlebnikov [91]	S	S Plebański [92] Stephani [93]	A

At a first glance, the progress made so far in this field looks quite satisfactory. However, with the exception of the plane waves (type N, $\varrho = 0$) and type D solutions our present knowledge is limited to solutions of comparatively little physical importance — or at least this importance has not yet been discovered.

In our opinion, the most interesting developments of the last years are connected with three questions.

The first question concerns the *twisting type N vacuum* solutions. It is known that the expanding non-twisting type N vacuum solutions must have singular lines. It is an open question whether twisting solutions may describe more realistic radiation fields. The only known solution was found by Hauser [82], but unfortunately it is not asymptotically flat [94].

The second problem concerns the *type D Einstein-Maxwell nonnull* fields. In the vacuum case, all type D solutions are known. Due to the efforts of Debever [95], Plebański-Demiański [83], and Leroy [84, 85], the problem to find all diverging type D Einstein-Maxwell fields seems to be almost settled.

Finally, progress has been made in finding new type N pure radiation fields [92].

Summary

To sum up: a lot of exact solutions are known. The storage of solutions which admit, to a certain extent, a physical interpretation has increased. Methods for treating non-linear partial differential equations, which were known in other branches of mathematical physics, have been successfully used in General Relativity.

Progress might result from new interpretations of already known solutions. Schmidt, in his abstract [96] for this conference, presented a nice idea: the Einstein-Rosen waves, usually interpreted as cylindrical waves, can be globally reinterpreted such that some of them describe radiation fields which become Minkowski space in the remote future.

Many problems are still unsolved. Is it possible to extend the generation methods for stationary axisymmetric vacuum fields to the perfect fluid case and to find an interior Kerr solution? Does a system with purely gravitational interaction emit radiation? In our opinion, this basic question cannot be answered convincingly until appropriate exact solutions will be available.

References

[0] Kinnersley, W.: in: Shaviv, G., and Rosen, J. (Ed.), General Relativity and Gravitation (Proceedings of GR7, Tel-Aviv 1974). — New York, London: Wiley 1975.
[1] Kramer, D.; Stephani, H.; MacCallum, M., Herlt, E.: Exact Solutions of Einstein's Field Equations. — Berlin: VEB Deutscher Verlag der Wissenschaften; Cambridge: Cambridge University Press 1980.
[2] Szafron, D. A.; Wainwright, J., J. math. Phys. 18 (1977) 1668.
[3] Ozsváth, I., J. math. Phys. 6 (1965) 590.
[4] Farnsworth, D. L.; Kerr, R. P., J. math. Phys. 7 (1966) 1625.

[5] Hiromoto, R. E.; Ozsváth, I., GRG **9** (1978) 299.
[6] Petrov, A. Z.: in: Recent Developments in General Relativity. Oxford: Pergamon Press and Warsaw: PWN 1962, 379.
[7] Griffiths, J. B., Proc. Cambridge philos. Soc. **77** (1975) 559.
[8] Bonnor, W. B., Phys. Letters **A 75** (1979) 25.
[9] Bertotti, B., Phys. Rev. **116** (1959) 1331.
[10] Robinson, I., Bull. Acad. Polon. Sci., Ser. Math. Astr. Phys. **7** (1959) 351.
[11] Bell, P., and Szekeres, P., Internat. J. theor. Phys. **6** (1972) 111.
[12] McLenaghan, R. G.; Tariq, N., J. math. Phys. **16** (1975) 11; **17** (1976) 2192.
[13] Tupper, B. O. J., GRG **7** (1976) 479.
[14] Ray, J. R.; Thompson, E. L., J. math. Phys. **16** (1975) 345.
[15] Coll, B., C. R. Acad. Sci. Paris **A 280** (1975) 1773.
[16] Kantowski, R.; Sachs, R. K., J. math. Phys. **7** (1966) 443.
[17] Collins, C. B., Commun. math. Phys. **23** (1971) 137.
[18] Dunn, K. A.; Tupper, B. O. J., Astrophys. J. **204** (1976) 322.
[19] Vajk, J. P.; Eltgroth, P. G., J. math. Phys. **11** (1970) 2212.
[20] Siklos, S. T. C., Preprint. Univ. of Oxford 1978.
[21] MacCallum, M. A. H.; Siklos, S. T. C., Abstr. GR9 1980.
[22] Evans, A. B., Mon. Not. Roy. Astron. Soc. **183** (1978) 727.
[23] Send, W.: Diplomarbeit, Hamburg 1972.
[24] Maartens, R.; Nel, S. D., Commun. math. Phys. **59** (1978) 273.
[25] Barrow, J. D., Nature **272** (1978) 211.
[26] Ruban, V. A.: Preprint. Leningrad 1978.
[27] Wainwright, J.; Ince, W. C. W.; Marshman, B. J., GRG **10** (1979) 259.
[28] Barnes, A., J. Phys. **A 11** (1978) 1303.
[29] Tariq, N.; Tupper, B. O. J., GRG **6** (1975) 345.
[30] Damião Soares, I.; Assad, M. J. D., Phys. Letters **A 66** (1978) 359.
[31] Dunn, K. A.; Tupper, B. O. J., Astrophys. J. **235** (1980) 307.
[32] Damião Soares, I., Rev. Bras. Fis. **8** (1978) 336.
[33] Kramer, D., J. Phys. **A 13** (1980) L 43.
[34] Tolman, R. C., Proc. Nat. Acad. Sci. U.S. **20** (1934) 69.
[35] Kustaanheimo, P.; Qvist, B., Comment. phys. math., Helsingf. **13** (1948) 12.
[36] McVittie, G. C.; Wiltshire, R. J., Internat. J. theor. Phys. **14** (1975); 145, **16** (1977) 121.
[37] Skripkin, V. A., Dokl. Akad. Nauk SSSR **135** (1960) 1072.
[38] Gutman, I. I.; Bespal'ko, R. M., Sbornik Sovrem. Probl. Grav., Tbilissi (1967) 201.
[39] Wesson, P. S., J. math. Phys. **19** (1978) 2283.
[40] Taub, A. H., Phys. Rev. **103** (1956) 454.
[41] Horský, J., Czech. J. Phys. **B 25** (1975) 1081.
[42] Tabensky, R. R.; Taub, A. H., Commun. math. Phys. **29** (1973) 61.
[43] Teixeira, A. F. da F.; Wolk, I.; Som, M. M., J. Phys. **A 10** (1977) 1679.
[44] Winicour, J., J. math. Phys. **16** (1975) 1806.
[45] van Stockum, W. J., Proc. Roy. Soc. Edinburgh **A 57** (1937) 135.
[46] Vishveshwara, C. V.; Winicour, J., J. math. Phys. **18** (1977) 1280.
[47] Wahlquist, H. D., Phys. Rev. **172** (1968) 1291.
[48] Szekeres, P., Commun. math. Phys. **41** (1975) 55.
[49] Szafron, D. A., J. math. Phys. **18** (1977) 1673.
[50] Tomimura, N., N. Cim. **B 42** (1977) 1.
[51] Szafron, D. A.; Wainwright, J., J. math. Phys. **18** (1977) 1668.
[52] Collins, C. B.; Szafron, D. A., J. math. Phys. **20** (1979) 2347.
[53] Bonnor, W. B., Commun. math. Phys. **51** (1976) 191.
[54] Bonnor, W. B.; Tomimura, N., Mon. Not. Roy. Astron. Soc. **175** (1976) 85.
[55] Neugebauer, G.; Kramer, D., Ann. Physik **24** (1969) 62.
[56] Kinnersley, W., J. math. Phys. **14** (1973) 651.
[57] Geroch, R., J. math. Phys. **12** (1971) 918.
[58] Ernst, J. F.; Wild, W. J., J. math. Phys. **17** (1976) 182.

[59] Kinnersley, W.; Chitre, D. M., J. math. Phys. **18** (1977) 1538; **19** (1978) 1926, 2037.
[60] Kinnersley, W.; Chitre D. M., Phys. Rev. Letters **40** (1978) 1608.
[61] Hoenselaers, C.; Kinnersley, W.; Xanthopoulos, B. C., Phys. Rev. Letters **42** (1979) 481; J. math. Phys. **20** (1979) 2530.
[62] Hauser, I.; Ernst, F. J., Phys. Rev. **D 20** (1979) 362; 1783.
[63] Hauser, I.; Ernst, F. J., J. math. Phys. **21** (1980) 1126; 1418.
[64] Neugebauer, G., J. Phys. **A 12** (1979) L 67.
[65] Neugebauer, G., J. Phys. **A 13** (1980) L 19.
[66] Harrison, B. K., Phys. Rev. Letters **41** (1978) 1197.
[67] Kramer, D.; Neugebauer, G., Phys. Letters **A 75** (1980) 259.
[68] Cosgrove, C. M.; J. math. Phys. **21** (1980) 2417.
[69] Belinsky, V. A.; Zakharov, V. E., Zh. Eksp. Teor. Fiz. **75** (1978) 1953; **77** (1979) 3.
[70] Wainwright, J., J. Phys. **A 12** (1977) 2015.
[71] Aleksejev, G. A., Abstr. GR 9 (1980) 2.
[72] Maison, D., Phys. Rev. Letters **41** (1978) 521; J. math. Phys. **20** (1979) 871.
[73] Neugebauer, G., J. Phys. **A 13** (1980) 1737.
[74] Harrison, B. K., Phys. Rev. **D 21** (1980) 1695.
[75] Kinnersley, W., J. math. Phys. **18** (1977) 1529.
[76] Wainwright, J., Commun. math. Phys. **17** (1970) 42.
[77] Wainwright, J., Internat. J. theoret. Phys. **10** (1974) 39.
[78] Oleson, M., J. math. Phys. **12** (1971) 666.
[79] Oleson, M., Ph. D. Thesis, University of Waterloo 1972.
[80] Robinson, I., GRG **6** (1975) 423.
[81] Held, A., N. Cim. Letters **11** (1974) 545.
[82] Hauser, I., Phys. Rev. Letters **33** (1974) 1112.
[83] Plebański, J.; Demiański, M., Ann. Phys. (USA) **98** (1976) 98.
[84] Leroy, J., Bull. Acad. Roy. Belg. Cl. Sci. **64** (1978) 130.
[85] Leroy, J., Bull. Acad. Roy. Belg. Cl. Sci. **62** (1976) 259.
[86] Kowalczyński, J. K.; Plebański, J. F., Internat. J. theor. Phys. **16** (1977) 371.
[87] Hacyan, S.; Plebański, J. F., Internat. J. theor. Phys. **14** (1975) 319.
[88] Debney, G., J. math. Phys. **15** (1974) 992.
[89] Herlt, E., GRG **12** (1980) 1.
[90] Stephani, H., J. Phys. **A 12** (1979) 1045.
[91] Frolov, V. P.; Khlebnikov, V. I., Preprint No. 27, Lebedev Phys. Inst. Akad. Nauk, Moscow 1975.
[92] Plebański, J., Private communication 1975.
[93] Stephani, H., Abstr. GR9 (1980) 76.
[94] Sommers, P.; Walker, M., J. Phys. **A 9** (1976) 357.
[95] Debever, R., Bull. Soc. Math. Belg. **23** (1971) 360.
[96] Schmidt, B. G., Abstr. GR9 (1980) 68.
[97] Herlt, E., GRG **11** (1979) 337.
[98] Kowalczyński, J. K., Phys. Letters **A 65** (1978) 269.

Address: Dr. D. Kramer and Dr. H. Stephani
Sektion Physik
Friedrich-Schiller-Universität, Max-Wien-Platz 1
DDR - 6900 Jena
German Democratic Republic

Computer Methods in General Relativity

R. A. d'Inverno (Southampton)

Introduction

High speed digital computers have been around for well over three decades, and yet it has only been in the last decade or so that computers have been used to any great degree in General Relativity. The area of application falls into the two distinct categories of numerical computing and algebraic computing. Much of the numerical work is of fairly recent origin and is still in development form, but nonetheless the literature is in a very accessible state thanks to the recent volume edited by L. Smarr [1] — indeed part of this review leans heavily on this reference and an excellent review article by T. Piran [2], where more detailed references may be found. Although algebraic computing seems to have been around longer its impact, outside the field of exact solutions, has been less obvious, and the associated literature is still in a rather patchy form (although see [3]).

In a review of this length we will only have an opportunity to attempt to convey some impression of the sorts of area of application involved, the methods employed and the results so far achieved. The hope is that it proves sufficient to give a flavour of the richness of the area by outlining the successes and the potential of computer methods. This is especially so since there is evidence of a rapid growth of interest in these techniques, and the consequent likelihood that their impact in the next decade will prove to be quite dramatic. The first part of the review is concerned with Numerical Methods and looks at the main method involving the 3 + 1 approach and the newer characteristic approach, and finally mentions briefly the Regge Calculus. The second part covers Algebraic Methods and after discussing some of the main ideas involved in algebraic computing, looks at a number of relativistic applications of the main systems and ends with a brief critique of the area.

1. Numerical methods

1.1. The 3 + 1 approach

Over half a dozen successful numerical codes have been developed for investigating spherical collapse, dust collapse, 2-dimensional axi-symmetric neutron star bounce, 2-dimensional two black hole collision, Brill waves, planar symmetry solutions, cylindrical symmetry solutions (for references see [1] and [2]) and colliding plane gravitational waves. In all but the last case, the technique involved has been based, directly or indirectly, on the 3 + 1 approach to General Relativity. We therefore start with a brief description of the 3 + 1 formalism, but for further details see, for example, the article of J. York in [1].

The 3 + 1 approach starts by slicing up space-time into a foliation of space-like hypersurfaces. The foliation is then rigged with a transvecting vector field. This vector field may be equivalently thought of as a fibration threading the foliation given by the congruence of curves to which the vector field is tangent. The foliation and fibration decompose the 4-metric $g_{\mu\nu}$ into the constituent parts

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu = \alpha^2\, dt^2 - g_{ij}(dx^i + \beta^i\, dt)\, (dx^j + \beta^j\, dt)$$

where g_{ij} (latin indices run from 1 to 3) is the 3-metric on the hypersurfaces, α is the lapse function which defines the foliation and β^i is the shift vector determining the fibration (see fig. 1). With this decomposition the ten Einstein field equations

$$G_{\mu\nu} = 8\pi T_{\mu\nu}$$

reduce to six evolution equations which are second order in time, and four constraint equations which are first order in time.

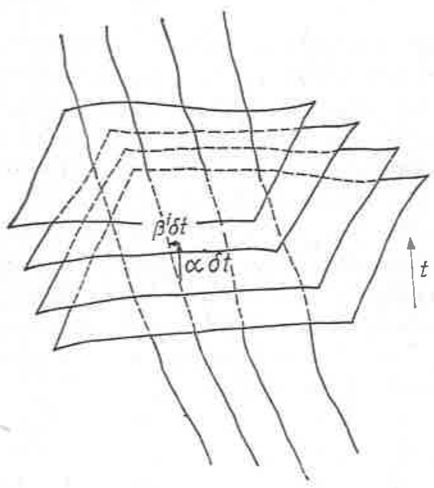


Fig. 1. Lapse and shift for a 3 + 1 foliation and fibration

In the usual formulation of the Cauchy IVP (initial value problem) one starts by prescribing data on an initial hypersurface, and then this data is propagated forward in time by means of the evolution equations. The initial data cannot be freely specified but must satisfy the constraint equations. However, if the constraints are satisfied on the initial hypersurface then, by virtue of the Bianchi identities, they remain satisfied for all future time. Thus the Cauchy problem is 'well-posed'.

In the majority of numerical schemes the six second order evolution equations are replaced by twelve first order equations. This involves introducing six additional variables, and it is now generally accepted that the natural choice of variables consists of the 3-metric g_{ij} together with the extrinsic curvature K^i_j (which is essentially the first time derivate of g_{ij}). The field equations then comprise twelve evolution equations

$$\frac{\partial}{\partial t} g_{ij} = -2\alpha g_{il} K^l_j + \beta^l g_{ij,l} + g_{il}\beta^l{}_{,j} + g_{jl}\beta^l{}_{,i}$$

$$\frac{\partial}{\partial t} K^j_i = \beta^l K^j_{i,l} + K^j_l \beta^l{}_{,i} - K^l_i \beta^j{}_{,l} - \alpha^j{}_{|i} + \alpha[K K^j_i + R^j_i + g^{lj} T_{il}]$$

and four constraint equations

$$R - K^i_j K^j_i + K^2 = 2\alpha^{-2}[T_{tt} - 2T_{ti}\beta^i + T_{ij}\beta^i\beta^j]$$

$$(K^i_j - \delta^i_j K)_{|i} = -\alpha^{-1}(T_{tj} - T_{jl}\beta^l)$$

where R^j_i is the 3-dimensional Ricci tensor and $_|$ denotes covariant derivative with respect to g_{ij}.

1.2. Numerical relativity

Smarr defines Numerical Relativity as a method of obtaining solutions of Einstein's field equations based on an evolution of initial value data defined on an edgeless spacelike hypersurface. The computational method for obtaining a solution is that of finite differencing. The procedure involves two separate steps: first the space-time is replaced by a finite lattice of grid points, and then the derivatives in the partial differential equations are replaced by a finite difference approximation. The exact solution to the finite difference equation depends separately on both the specification of the grid and of the finite difference approximation used. Unfortunately there are an infinite number of possible finite difference analogues, each with its own solution, but a large number of them will bear little resemblence to the exact solution of the original equations. This is because of instabilities which arise due to an incorrect choice of discretization of space-time. Even if one is using a stable scheme, another major source of inaccuracy occurs in truncation errors. These latter errors stem from the fact that one is essentially approximating a function by a finite part of a Taylor series expansion.

Now in the continuum case, initial data satisfying the constraint equations is evolved into data which, by virtue of the Bianchi identities, still satisfies the constraints. However, this relation breaks down in a finite difference scheme. In such a case the equations are written on a space and time lattice and due to the non-commutative nature of finite difference derivatives all the equations cannot be satisfied simultaneously in a trivial way (for example to the same order in space and time increments). The fact that the finite difference version of Einstein's equations leads to an overdetermined system is the most important current problem in numerical relativity. As a result one of two strategies have to be adopted, either the constraints are simply ignored which is termed free evolution, or they have to be artificially imposed. In the latter case, one method involves imposing the constraints after finite intervals of time (chopped evolution) and another is to impose them at every stage of the integration (fully constrained solution). Unfortunately, each method has an associated drawback. For example, computations with particular solutions have demonstrated that a freely evolved solution drifts further away from the true solution as it is evolved in time [4]. Similar problems arise with chopped evolution and fully constrained solutions. Piran has indicated this schematically in fig. 2, where the plane represents the subspace of solutions which satisfy the constraint equations [3].

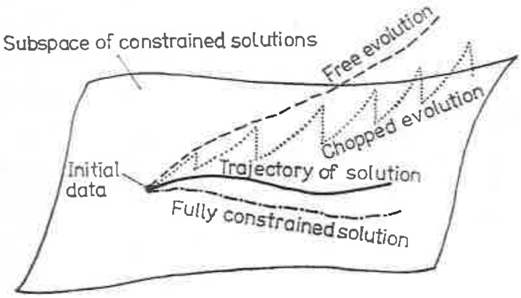


Fig. 2. Schematic representation of drift away of different numerical solutions from true solution

Other integration schemes have been proposed which involve the addition of auxilliary variables or auxilliary equations. but they all have the same attendant defficiencies. The empirical evidence suggests that a fully constrained solution is perhaps the best that can be currently obtained.

1.3. Coordinate and boundary conditions

General Relativity is a gauge theory in that it admits the 4-dimensional pseudo-group of co-ordinate transformations. In Numerical Relativity the exploitation of this fact by adopting a particular choice of coordinate (gauge) conditions is an important and integral part of any scheme. This usually takes the form of specifying the lapse and the shift directly, or of specifying conditions giving rise to equations which in turn determine the lapse and the shift. Now the choice of the lapse is particularly important because the resulting foliation determines what part of space-time can be explored. The choice of the shift can be thought of as determining the 3-dimensional coordinates on the slices.

The simplest choice of setting $\alpha = 1$, $\beta^i = 0$ (which is still used in several codes) does not work very well in general because the congruence emanating from the initial slice converges and eventually focussing occurs. Thus coordinate singularities result and so prevent a maximal development of the solution. As Lichnerowicz originally pointed out, this focussing can often be avoided if the foliation consists of maximal slices, that is those for which trace $K^i_j = 0$. Indeed it is now generally accepted that the best choice for the lapse is maximal slicing for open universes and constant mean curvature (trace $K^i_j = f(t)$) for closed (spatially compact) universes. However there are a number of open questions associated with this choice. For example, do maximal (constant) slices exist generally? Clearly the important question is to determine a slicing which allows one to obtain the maximal development. In the words of York [2] "the development of a constructive prescription for generating a foliation covering the maximal Cauchy development is the number one problem in the dynamical approach to the solution of Einstein's equations". Another conjecture concerns whether or not maximal slicing avoids intrinsic singularities. The conjecture pictures a maximal foliation wrapping itself around the singularity with the lapse collapsing to zero as the singularity is reached. There are counter-examples to this conjecture in non-vacuum space-times, but no known ones for vacuum space-times. As far as the shift is concerned there are a number of different conditions apart from setting $\beta^i = 0$ which have been adopted including a choice which diagonalises the metric and a rather complicated condition known as the minimal distortion shift vector condition [1]. It is probably true to say that all of these choices have difficulties and limitations associated with them.

Another problem area arises with boundary conditions. If initial data is prescribed on a finite portion S_0 of an initial slice, then the solution is only determined within $D^+(S_0)$ the Cauchy development of S_0. Since a numerical grid must of necessity be finite this limits the size of the development and so one may not be able to generate a solution which covers the whole area of interest. However if S_0 is a subset of a global Cauchy hypersurface, and the space-time is empty and asymptotically flat then the solution can be extended by specifying boundary conditions on $\mathbb{R} \times S_0$. This is limited by the attendant danger that false boundary conditions may change the solution within $\mathbb{R} \times S_0$ and so ultimately lead to a false solution. For example, in a collapse problem certain boundary conditions may enable one to extend the calculation as long as nothing crosses the boundary, but they reflect all outgoing perturbations (gravitational wave pulses) back into the solution. There have been techniques developed for overcoming this problem [4]. Another difficulty relates to inner symmetry boundaries (e.g. the axis of an axi-symmetric solution) where there are severe problems with existing codes in keeping the solution regular along such boundaries.

1.4. Computer calculation

To summarise, a numerical calculation typically breaks down into the following steps:

(i) choice of equations
(ii) choice of coordinate conditions
(iii) specification of initial data

(iv) solution of the initial value problem
(v) specification of boundary values
(vi) choice of grid lattice
(vii) transformation of equations into finite difference form
(viii) implementation on computer
(ix) interpretation of the results.

A considerable number of problems confront Numerical Relativity at present including numerical instabilities, truncation errors, coordinate conditions, coordinate singularities, limitations in computer time and memory and the difficulty of representing the solution from the final results. All the existing codes are essentially one or two dimensional in character. However three dimensional codes are beginning to become feasible, even though they entail enormous demands on computer time and store. Another problem which is also causing difficulties involves the successful treatment of relativistic shock waves, although some advances have recently been reported [5, 6, 7].

1.5. The characteristic approach

The $3 + 1$ approach fails if the foliation becomes null, because the metric g_{ij} of a null hypersurface is degenerate. Yet characteristic initial value problems are of interest in their own right for a number of reasons. First of all they present a natural vehicle for studying gravitational radiation because information propagates along null geodesics which rule null hypersurfaces. Again a Bondi-type problem involving studying the asymptotics of isolated radiative systems gives rise to a characteristic IVP. Another obvious candidate is Cosmology where, after all, we gather information about the Universe by observations along our past null cone. From a calculational viewpoint one main advantage of this approach is that hyperbolic partial differential equations reduce to ordinary differential equations along characteristic curves.

An attempt to develop a formalism for characteristic IVP's analogous to the $3 + 1$ formalism was begun by the author and J. Stachel [8] and completed more recently with J. Smallwood [9]. The resulting $2 + 2$ formalism begins by foliating space-time into two families of space-like 2-surfaces. Each family can then in turn be considered as foliating a three-dimensional hypersurface which may either be null, timelike or spacelike depending on the IVP under investigation. Figure 3 illustrates schematically two of these hypersurfaces for a double null and null-timelike IVP. The $2 + 2$ formalism then identifies the so-called conformal 2-structure — essentially the family of conformal metrics of the 2-surfaces — as dynamical degrees of freedom. The Einstein field equations involve only two evolution equations propagating the conformal 2-structure. This formulation avoids the difficulty of the $3 + 1$ approach since the initial data is no longer constrained. Formal integration schemes have been obtained for double-null, null-timelike, null-spacelike and Cauchy (spacelike-timelike) IVP's. The formalism has the property of not only being manifestly covariant, but also of clarifying the geometrical significance of the gauge freedom which exists. It is hoped that this approach may provide a possible route towards quantisation since, unlike the 3 + 1 approach, the constraints are fully eliminated. No "hard" theorems for existence and uniqueness have yet been obtained in general for all these IVP's, and it may be that in certain cases (for example those involving spacelike hypersurfaces) the problems may not be well-posed in anything but the rather restrictive case of analytic solutions. However, such theorems have been obtained for the double-null IVP by H. Müller zum Hagen and H. Seiffert in the harmonic gauge [10] and by J. Stewart in the Newman-Penrose gauge [11].

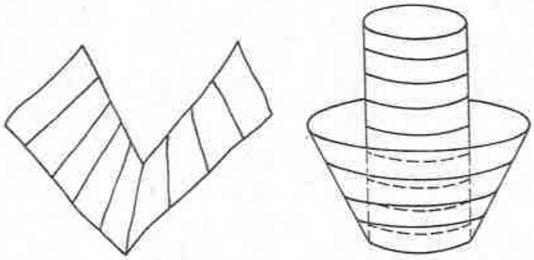


Fig. 3. Spacelike 2-surface foliation for two hypersurfaces in the double null and null-timelike IVP

Only two numerical codes have been developed to date which involve the characteristic approach. G. Bicknell and R. Henriksen have a code for spherical collapse [12], evolving their solution along the hydrodynamic characteristics (no information propagates along the null geodesics because of the spherical symmetry). More excitingly, J. Stewart and R. Corkhill have used the Newman-Penrose formalism to handle the double-null characteristic IVP [11]. They have applied this work in particular to investigating colliding plane gravitational waves. Their resulting numerical code has been able to follow the evolution of the waves up to the formation of the intrinsic singularity (see fig. 4). Using recent results of H. Friedrich [13], it is possible to show that the constraints and Bianchi identities are always satisfied (to within the usual numerical accuracy) in this method. These same workers are also exploiting the Penrose technique of conformal compactification to consider asymptotic characteristic IVP's (see for example [14]). There seems little doubt that characteristic codes are likely to provide a very fruitful method for obtaining numerical solutions in the near future.

1.6. Regge calculus

There are two principal ways of solving partial differential equations on a computer. The first involves finite differencing where one essentially discretises the derivatives occurring in the equations, and the other involves a finite elements method where one essentially discretises the underlying space. The Regge Calculus is related to the class of finite element methods and has attracted attention in recent years as

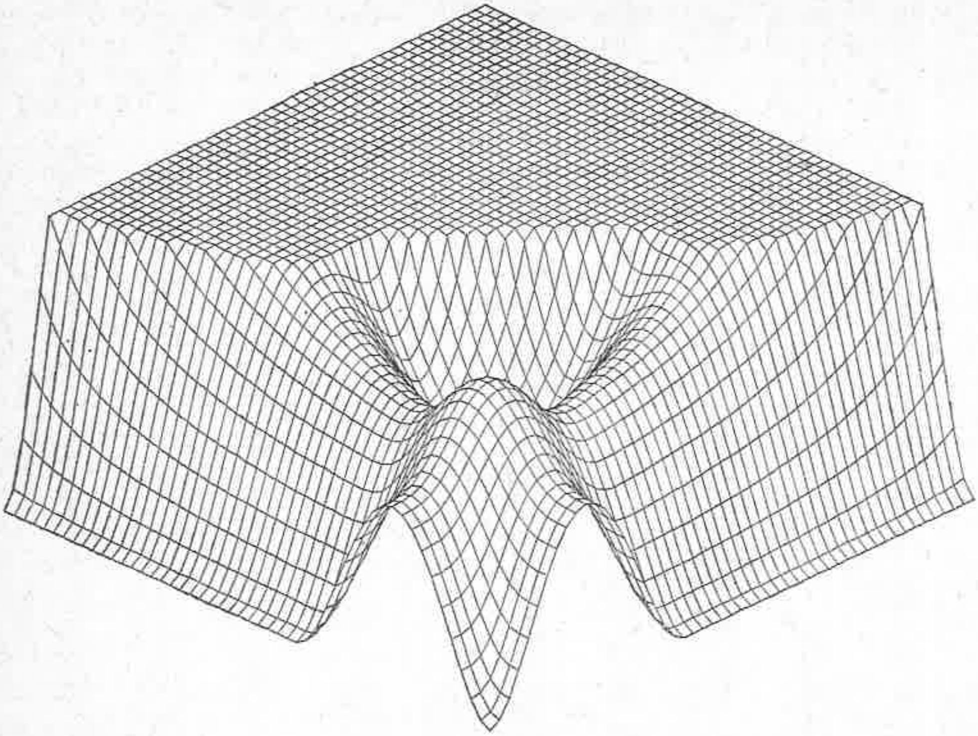


Fig. 4. Two colliding plane gravitational waves (chopped off as curvature mounts towards the singularity)

a numerical method for solving Einstein's equations [15, 16, 17]. In Regge Calculus, one replaces the four-dimensional continuous curved space-time with four-dimensional flat simplices. The curvature is concentrated in a δ-function form on the bones or two-dimensional edges of these simplices and can be expressed in terms of a deficit angle associated with each bone.

The basic advantage of this approach is that it no longer leads to an overdetermined system, since there is a one-to-one correspondence between the variables (the lengths of the legs or edges of the bones) and the equations they satisfy. However the equations do not separate into evolution and constraint equations but are rather on the same footing. This presents difficulties in formulating an IVP. In addition the construction of the simplices is rather an elaborate process and the resulting equations are complicated. Although there is some debate, the consensus would seem to hold that finite element methods are better suited to solving elliptic equations and finite difference methods to hyperbolic equations. Indeed, examples are known of hyperbolic equations where the standard finite elements procedure breaks down. However there are techniques which essentially transform hyperbolic equations to elliptic ones (by treating the time coordinate separately) and then finite element methods can be used successfully. It would be very helpful to know to what extent these difficulties with finite element methods apply to Regge Calculus. In any case, Einstein's equations become elliptic for spherically symmetric or static configurations and Regge Calculus has been shown to give convergent solutions in these cases [15, 16]. The time development of some simple model universes has also been computed from Regge Calculus in a type of 3 + 1 approach where the time is treated essentially as a continuous variable [17]. Once a Regge Calculus space-time has been constructed, then it should be possible using a recently reported technique [18] to compute the space-time geodesics.

2. Algebraic methods

2.1. Background

Algebraic computing came into existence because, in the words of Jean Sammet, "... it has become obvious that there are a large number of problems requiring very tedious time-consuming error-prone and straightforward algebraic manipulation and these characteristics make computer solution both necessary and desirable" [19]. A. C. Hearn has pointed out that one central area where algebraic computations arise in theoretical physics is in the application of a theoretical model to some 'real' problem [20]. It is rarely the case that the problem can be treated exactly and so recourse has to be made to approximation techniques. The methods involved, such as perturbational analysis, are better suited to the exact arithmetic of algebra rather than the less accurate numerical methods which are currently employed. To date, significant applications of algebraic computing have been made in the fields of Quantum Electrodynamics, Quantum Mechanics, Celestial Mechanics, Fluid Mechanics and, increasingly, in General Relativity.

In the case of relativity the most frequent application has been the so-called metric application. This involves calculating curvature tensors and related quantities from a given metric form. Such a calculation is clearly algorithmic since it can be broken down into a series of well-defined steps (which is a necessary prerequisite for a computer calculation). A specific example, which has become a canonical one for attempting comparisons between different algebra systems [3], is the Bondi metric used for studying isolated radiative systems, namely

$$\mathrm{d}s^2 = (Vr^{-1}\,\mathrm{e}^{2\beta} - U^2r^2\,\mathrm{e}^{2\gamma})\,\mathrm{d}u^2 + 2\mathrm{e}^{2\beta}\,\mathrm{d}u\,\mathrm{d}r + 2Ur^2\,\mathrm{e}^{2\gamma}\,\mathrm{d}u\,\mathrm{d}\theta$$
$$-r^2\,\mathrm{e}^{2\gamma}\,\mathrm{d}\theta^2 - r^2\,\mathrm{e}^{-2\gamma}\sin^2\theta\,\mathrm{d}\varphi^2$$

where V, U, β and γ are all functions of three of the coordinates u, r and θ. Figure 5 illustrates an excerpt from a SHEEP program in which the Bondi metric is read in from SHEEP's library of solutions and one of the components of the Ricci tensor is computed. In the original hand calculation of the Ricci tensor the work was spread over something like a six month period. Nowadays the more efficient algebra systems can duplicate the calculation in a little over ten seconds. Moreover, unlike the original hand calculation, the computer results are error free. Before considering some of the issues involved in algebraic computing we shall introduce a little of the computer terminology involved.

. . .

*(DSKIN BONDI)

Bondi radiating metric
Bondi et al 1962 Proc. Roy. Soc. A, vol. 269, p. 21

*(WMAKE DS2)

$$ds^2 = \left((Vr^{-1} e^{2B} - r^2 e^{2G} U^2)dt^2 + 2e^{2B} \; dtdr + 2r^2 e^{2G} \; UdtdH\right) - r^2 e^{2G} dH^2 - r^2 e^{-2G} \; \sin^2(H)dF^2$$

*(WMAKE RIC11)

$$R_{11} = -2G_{,r}^{2} + 4r^{-1} B_{,r}$$

. . .

Fig. 5. SHEEP interaction excerpt (typed input is underlined)

2.2. Computer terminology

The modern computer in its barest essentials is illustrated in fig. 6. First of all there is a medium for communication with the computer. The information which the computer is fed, termed input, is often in the form of punched cards and then the corresponding information received back, the output, consists of sheets of typed material known as computer listing. Increasingly though the computer terminal is becoming the most frequently used form of communication. It consists of a typewriter keyboard linked to some form of display, either on a video screen or in hard print form. These terminals enable users to communicate very much more readily with the computer. Moreover they may be quite remote from the actual site of the computer, simply being connected to it by some form of telephone link-up. The pieces of equipment which handle the different sorts of input and output are known as peripheral devices. The computer proper mainly consists of a device for storing large amounts of information, usually in the form of numbers or characters. This is known as the memory, the core or the main store of the computer. The actual manipulation of this information is then carried out by the 'brain' of the computer called the central processor. Finally the storage capabilities are greatly extended by the so-called backing store which usually consists of disks or drums and magnetic tapes. Main store differs from backing store in that information in it can be accessed directly and very rapidly whereas information in the backing store has first to be loaded into the main store before it can be utilised. However, since the backing store is relatively very large, users may reserve areas of it for permanent storage. In the case of disks these areas are known as files. The advent of terminals for communication and files for storage in recent years has significantly eased the activity of computing. The work presented to the computer for processing is termed a job. In older machines these jobs are batch processed which means that they queue up and are processed one by one. Most modern machines are time sharing which means that they are capable of processing a number of jobs simultaneously (this is possible, for example, since one job may be processed whilst input is awaited from the various terminals hooked up to the machine).

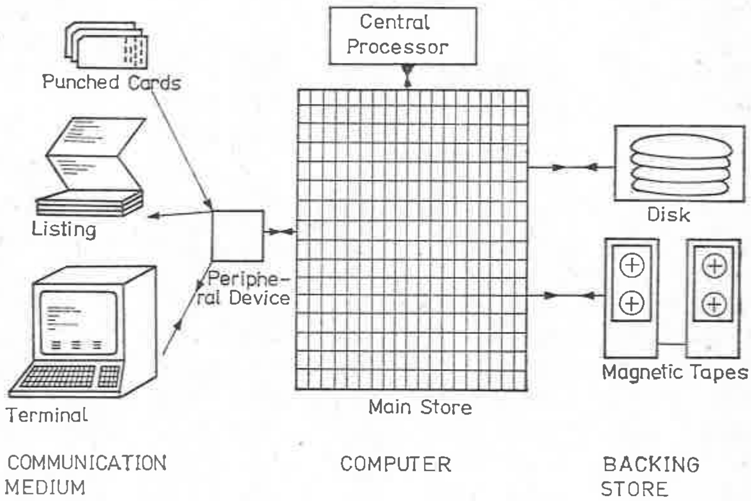


Fig. 6. Skeletal Form of Computer

The actual instructions for the machine to carry out are contained in the program. This is written in a special language which can vary quite markedly in appearance. In a low level programming language the instructions are of a very detailed nature involving numerical codes which vary from one make of computer to another. At the other extreme, high level languages are much closer to English language statements and to mathematical notation and therefore are significantly easier to write in. More importantly they are machine-independent and so work on all machines possessing the language. Computing is perhaps at its most efficient when a programming language is an interactive one. This means that the language possesses a conversational component so that it is possible to communicate directly with the computer and obtain responses which are essentially spontaneous. Unfortunately most languages do not as yet have interactive versions generally available. The best known high level languages are FORTRAN and ALGOL and they are mainly used for numerical work. Although many algebra systems are FORTRAN or ALGOL based, the majority of the most successful ones are written in a rather esoteric language called LISP.

Because of its importance in algebra we shall discuss two of the central properties of LISP.

LISP offers a natural representation for algebraic expressions. This is because the data items of LISP are objects called lists. A list is essentially a row vector whose elements may in turn be row vectors. For example, consider the mathematical expression

$$a + bc.$$

First of all we need to make all operators explicit, so that if we denote multiplication by the symbol $*$, we can rewrite this as

$$a + b * c.$$

The next step is to change this infix notation to a prefix form by writing each operator first followed by its arguments. Then this expression becomes

$$\big(+a(* \, bc)\big).$$

This is now a list. It consists of three items, the first two being $+$ and a. The last item is itself a list consisting of the three items $*$, b and c. Figure 7 gives a representation of how a structure like this is actually stored in the computer. In a word-orientated machine the storage is divided up into words which are usually used to store a number. Lists are formed by chaining a sequence of words together. This is done by dividing the word into two halves, the first half containing a representation of the corresponding item in the list and the second half containing the address in store of the next word in the chain. LISP is a language for constructing, searching and modifying lists.

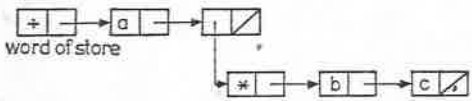


Fig. 7. Internal representation of the LISP expression $(+a(* \, bc))$

Now the main way in which an algebraic computation is different from a numerical one is the existence of intermediate term swell. Most algebraic calculations involve expanding out expressions as a preliminary to cancellation, and these intermediate expressions are often very large. In a complicated calculation they could swell to fill up the whole machine — known as 'blow-up' — and so cause the program to halt through lack of store. However many of these expressions are no longer needed since they have arisen through intermediate steps in a calculation. If we could throw these redundant expressions away, then we could reclaim the store they occupy and so the program could start up again. In fact LISP possesses an automatic device called a garbage collector which does precisely this. The ability of a program to start itself up again when it runs out of store proves to be an essential prerequisite in most cases for the construction of an efficient algebraic computing system.

2.3. Issues in algebraic computing

We shall outline some of the main issues which confront designers of algebra systems. The first relates to the problem of simplification. Simplification is needed in order to keep expressions compact both to economise on store and to render them intelligible to the user, and to check whether or not an expression is zero. The first problem is that simplification is not in general algorithmic since there does not exist a unique form for a fully simplified expression. However the problem is more profound because Richardson has proved that simplification is not even decidable [21]. That is, it is not possible to decide in a finite number of steps whether or not an algebraic expression of sufficient richness is zero (specifically he considered that class of expressions built out of the integers, π and x by means of addition, multiplication, division, substitution, sine, exponential and logarithm). This places an ultimate restriction on the capabilities of an algebra system. Nonetheless, ad-hoc procedures have been developed which work extremely well in practise.

The next issue concerns the factorisation of polynomials. A simple example occurs in the simplification

$$\frac{x^2 - y^2}{x + y} \to \frac{(x + y)(x - y)}{x + y} \to x - y.$$

Several systems possess factorisation capabilities of some form or other, but although sophisticated factorisation programs are now available they tend to be very large specialised systems and any problem involving large amounts of simplification would need to make use of such a system. The design of present day operating systems controlling the organisation of a computer make transfer of information between different systems a relatively straightforward procedure in principle. One of the biggest breakthroughs in recent years concerns the integration of elementary functions. Originally integration, in contradistinction to differentiation, was not considered to be an algorithmic process. However Risch has discovered a decidable procedure which can determine whether or not an expression is integrable within a given class of functions and, when it is, gives a constructure prescription for determining the integral [22]. This has recently been implemented on the system REDUCE and partly on the system MACSYMA. The integration programs are very large and should again be considered as specialist systems.

The biggest remaining issue concerns the question of user orientation. One aspect of this relates to the format of mathematical expressions. Most of the systems possess a FORTRAN or ALGOL like infix notation for the input of expressions, and it does not take very long to learn how to use it. In an interactive environment input expressions can be echoed back for instantaneous checking. A fair proportion of the systems possess a two-dimensional format for displaying output expressions, thus allowing superscripts, subscripts, powers and the like to be easily recognised. One advantage is that these formats are immediately intelligible to the non-specialist. One of the present limitations involves the chracter-set available, which usually consists of upper or lower case latin letters only. Other aspects of user orientation revolve around the ease, or otherwise, with which a new user can become an actual user of a system. CLAM claimed to be a system which could enable a new user to process most metrics after about half an hour's study. SHEEP has a fairly readable manual which starts off with some demonstrations of interactive sessions processing particular solutions. Most manuals are unfortunately notoriously difficult to use. Interactive languages may be helpful in this respect, since they essentially enable a user to de-bug a program as it is being written. A suggested direction for improvement is contained in the design of the MACSYMA primer. This is an interactive program which aims to teach a new user how to use MACSYMA step by step, offering choices for items to be studied, exercises to attempt, and the like. Designing systems in such a way that they make the conversion of a potential user into an actual user in as easy a manner as possible is an important problem which has often been overlooked in the past when efforts were concentrated on making systems more powerful and efficient.

Table

	LISP-based	FORTRAN-based	Other
General purpose	MACSYMA REDUCE SCRATCHPAD ...	FORMAC† ALPAK† ALTRAN SAC-1 METASAC GOEDEL [24] ...	CAMAL SCHOONSHIP SYMBAL ...
Designed for G. R.	LAM ALAM† ILAM† CLAM† SHEEP ORTOCARTAN [23] ...	GRATOS [25] ...	...

† Largely obsolete

2.4. Algebraic systems

The following table contains the more important systems, as far as relativity applications are concerned, together with three recently reported systems. In the opinion of the author, the 'big four' are SHEEP, CAMAL, MACSYMA and REDUCE. SHEEP and MACSYMA are interactive whereas most versions of CAMAL and REDUCE are not (this is perhaps less of a disadvantage with CAMAL which has a rather different design philosophy). SHEEP was written especially for relativity, particularly with metric applications in mind. The other three are general algebraic systems but possess relativity packages. Further details and references may be found in [3] and [26]. In the next section we shall consider examples of relativity applications of three of the big four, together with the systems LAM, ALAM and CLAM which although they have been substantially used are largely now of historic interest.

2.5. Some relativistic applications

2.5.1. LAM, ALAM, CLAM

Descriptions of these systems may be found in [27, 28]. They are now essentially obsolete, having been superseded by the system SHEEP. However, among the applications which they have been used for we mention the following:

(i) Well over 100 metrics have been processed by them. Included in these is the classification of the 40 vacuum solutions due to B. Harrison [29] (in fact 4 turned out to be non-vacuum [30].) Many of these metrics are exceedingly complicated and it has been estimated that if the calculations had been carried out by hand it would represent more than a life-time's work.

(ii) The systems had limited capabilities for calculating frame components of certain quantities. These were helpful, for example, in finding the principal density and pressures of a given energy-momentum tensor.

(iii) These were the first systems to process a program for determining the Petrov type of a metric [30]. This is clearly an algorithmic calculation since it essentially involves classifying the roots of a quartic equation.

(iv) The system could perform functional differentiation. This facility was used for variational calculus investigations. In particular, the Noether identity for Bondi's metric was investigated with this tool and it eventually led to a new formalism [31, 32] which was a precursor to the $2+2$ formalism reported earlier in this review.

(v) ALAM had some very sophisticated capabilities for handling truncated power series expansions. They were used, in particular, for obtaining a power series expansion of the vacuum Bondi solution in inverse powers of the luminosity distance parameter (which was taken to the eighth order compared to the second order previously obtained by hand), and producing an independent derivation of the Bondi massloss result by an asymptotic investigation of the Landau-Lifschitz pseudotensor.

(vi) At GR 7 in a review of exact solutions, W. Kinnersley reported that there were no known twisting type N or twisting type III vacuum solutions. CLAM was used to check the resulting discoveries of I. Hauser [33] for twisting type N, and I. Robinson [34] and A. Held [35] for twisting type III. This investigation led to the new technique of functional form invariance which may be of use, among other things, for obtaining killing solutions [36, 37].

2.5.2. SHEEP

A description of this system is given in [38], and an illustration of the power and versatility of the interactivity of SHEEP may be found in [39]. There are now about ten institutions which have access to SHEEP, and there are consequently a significant number of applications of the system. Of those known to the author we mention the following:

(i) There is a coordinate component version and a frame component version of SHEEP. The frames may be chosen to be null, quasi-null or Lorentzian in particular.

(ii) There is a growing library of most of the well known solutions (in both coordinate and frame component form).

(iii) There is a Petrov classification package and an Einstein-Maxwell package.

(iv) SHEEP possesses truncated power series facilities. They have been used by J. Åman [40] to check and extend the work of J. Synge and P. Florides on rotating bodies in General Relativity [41].

(v) It is relatively easy to write programs in SHEEP to compute new tensors. One application has involved calculating the differential invariant $R_{\mu\nu\varrho\sigma;\tau}R^{\mu\nu\varrho\sigma;\tau}$ for black hole solutions which has led to a prescription for avoiding black holes [42].

(vi) Perhaps the most exciting recent development relates to the local equivalence problem in General Relativity, that is the problem of deciding whether or not, given two metrics $g_{\mu\nu}$ and $\bar{g}_{\mu\nu}$, there exists a coordinate transformation which transforms one metric into the other. Following on the original work of C. Brans [43], A. Karlhede has now developed an algorithm which makes computer calculation a viable proposition [44]. The algorithm involves calculating at most the seventh covariant derivative of the Riemann tensor for type D or N metrics and the fifth covariant derivative for types I, II or III. However this degree is reduced if the metric possesses an isometry group or if there are functional relationships between the components of the Riemann tensor in a particular frame. The algorithm has been implemented in SHEEP by J. Åman [45]. It turns out in practise that only the first covariant derivative has been required for those solutions investigated so far (see (4) (iii) below in this connection). As an example of a new result obtained with this program, they have found that the Harrison solutions III — 9(a) and III — 9(b), which were previously considered distinct, are in fact the same [45]. This suggests that there may be slight defficiencies in W. Kinnersley's classification scheme for type D solutions [46]. The ultimate objective of this work is to compile a library of *all* known solutions on computer file. Moreover, these solutions would be fully classified. Then any 'newly' discovered solution could be checked for originality by the equivalence program, and the library updated accordingly. This would then hopefully avoid duplicate discoveries, as well as providing an easily accessible catalogue of exact solutions.

(vii) Finally we mention the very sophisticated system STENSR of L. Hörnfeldt [47]. This was originally implemented in SHEEP but has also recently been implemented in MACSYMA as well. STENSR is a tool for carrying out tensor or spinor calculus with symbolic indices. It can cope with covariant differentiation, complicated symmetries, contractions, and complex conjugation among other things. It can also exploit side relationships, such as $\sin^2 + \cos^2 = 1$, in a fairly optimal way. This turns out to be crucial in certain calculations resulting in the collapse of enormous expressions to a manageable size. In addition the system possesses a 'tensor compiler' which enables one to define new tensorial or spinorial quantities in terms of existing ones using formulae very similar in character to hand written formulae. The program can then automatically calculate the components of these quantities for a particular solution, thus avoiding the need to write a particular program in SHEEP (or LISP) to accomplish this task. STENSR has a growing number of applications in classical relativity, quantum gravity and supergravity [47, 48].

2.5.3. MACSYMA

A description of this system may be found in [49], and some of its applications in General Relativity are discussed in [50]. In particular we mention: —

(i) MACSYMA has been used to investigate the viability of gravity theories alternative to General Relativity [50]. In particular R. Pavelle found that Yang's theory (with 'pure space' equations $R_{\mu\nu;\sigma} - R_{\mu\sigma;\nu} = 0$) admits a static spherically symmetric solution additional to that of Schwarzschild. He then demonstrated that this property violates Birkhoff's theorem. (In fact MACSYMA was used to clarify the status of Birkhoff's theorem for a non-vacuum solution showing that the energy-momentum tensor must be diagonalisable and static). Similar problems were found with Brans-Dicke theory and Yilmaz theory. In addition Pavelle has found that the theory of Mansouri and Chang is definitely inequivalent to General Relativity.

(ii) MACSYMA possesses a package called ITENSR for performing indicial tensor manipulation. The ability of this package to carry out covariant differentiation led to the discovery of a new algorithm for performing multiple covariant differentiation [51]. This package (together with STENSR) has also been used to investigate the Riemann invariant densities [48]

$$K(m) = \sqrt{g}\, \delta^{\mu_1\mu_2\cdots\mu_m}_{\nu_1\nu_2\cdots\nu_m} R^{\mu_1\mu_2\cdots\mu_{2m-2}}_{\nu_1\nu_2\cdots\nu_{2m-2}} R^{\mu_{2m-1}\cdots\mu_{2m}}_{\nu_{2m-1}\cdots\nu_{2m}}.$$

In particular

$$K(2) = R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\varrho\sigma}R^{\mu\nu\varrho\sigma}$$

was confirmed as satisfying the Gauss-Bonnet theorem in four dimensions (which has importance in both quantum gravity and supergravity). In addition $K(3)$, $K(4)$ and $K(5)$ have been computed explicitly. If fully expanded by hand $K(5)$ would involve 3.6 million terms and would take an estimated five man years to complete. After simplification it reduces to 93 quintic terms and takes about 6 hours to compute by machine.

2.5.4. CAMAL

There is a manual describing some of the system's capabilities [52], and a relativity package which has been developed for CAMAL [53]. We mention the particular applications: —

(i) CAMAL was used by J. Fitch to check many of the calculations originally carried out by ALAM and CLAM, including the classification of the Harrison metrics and the expansion of the Bondi vacuum solution [54]. Included among the exact solutions investigated by CAMAL is the Sato-Tomimatsu solution which was identified as a type I naked singularity [55]. In addition some inhomogeneous cosmological models due to P. Szekeres have been classified [56]. These have been extended to obtain a new class of inhomogeneous and anisotropic cosmologies with perfect fluid matter content [57]. It has also been used to identify some new type D solutions with perfect fluid sources [58].

(ii) J. Wainwright and S. Campbell were the first to implement the Newman-Penrose formalism on a computer [59]. In particular it turns out to be more efficient to process certain metrics by starting from a null tetrad.

(iii) M. MacCallum has also embarked on a similar project to that previously reported concerning the equivalence problem. He hopes to use CAMAL to set up a library of all known solutions [60]. The recently published volume of exact solutions will provide an excellent starting point for this venture [61]. Moreover, MacCallum intends to exploit an algorithm for the equivalence problem due to S. Siklos [62], where it is believed that, in all cases, it may only be necessary to compute the first covariant derivative of the Riemann tensor (or its equivalent). If this turns out to be the case then it should significantly improve the chances of a successful outcome for this important project.

2.6. Advantages and disadvantages

There would appear to be three different ways in which algebra systems can be employed. First of all they can be used as a desk calculator for spontaneous help in solving day-to-day problems. (In fact a desk calculator algebra system already exists and it may not be too fanciful to anticipate the future existence of pocket algebra calculators). Next, they are clearly of great help in solving straightforward (that is algorithmic or semi-algorithmic) but very tedious problems. Finally they make certain calculations possible which would otherwise be considered intractable because the time taken to complete them by hand would be prohibitive. We have already pointed out that the whole area of perturbational analysis will profit from the existence of these systems. Perhaps the other point to emphasise is that the results obtained from a computer calculation are much more likely to be error free. There may be some difficulty in accepting this fact by those who are unfamiliar with the internal workings of these systems. However a sufficiently large number of complicated calculations have been carried by systems designed in such completely different manners and yet all giving identical results, that they should command as much faith as is given to numerically-based computer calculations.

The experience of the author is that the best vehicle for algebraic computing consists of an efficient system which is interactive. In most applications, for example, substitutions play an important rôle and it is rarely clear beforehand exactly what these substitutions should be. In such a case man-machine interaction would seem to be the best expedient (see [39] in this connection). Finally we mention the fact that it has happened on more than one occasion that a very complicated calculation, perhaps involving thousands of intermediate terms, has ended up producing a very simple result. This then suggests that there may be some underlying structure, and a closer investigation of the problem has then led to a means of reproducing the result which is independent of any intermediate computation. This is particularly significant when one takes into account the fact that the problems may not have otherwise been tackled because of the apparent complexity of the calculations involved.

The biggest current drawback with algebra systems is that they are all more or less machine-dependent. For example, SHEEP only operates on a DEC PDP 10 computer (although there are plans to write a version in canonical LISP which would make it available on considerably more machines). This machine-dependence is less of a problem in the U.S. where most of the big computers are linked together through a national network enabling users in one installation access to computers in other installations. One answer to this drawback then would be the introduction of more and bigger networks. Other systems, such as MACSYMA, are not for general distribution and so again cannot be accessed by most users. Another limitation lies in the fact that many of the systems are not interactive. It is also probably true to say that although system documentation is improving, it still is generally rather poor and this tends to deter potential users. As we have already pointed out, self-instructional packages would help in the initial stages. This should then be allied to some form of system documentation available on file which could be looked up for help during a computing session.

Despite these defficiencies one can nonetheless point to a fairly impressive list of succesful algebraic computations. As there must be something approaching a hundred users in existence, there is likely to be a substantial increase in these applications in coming years. Their impact may not be quite so easy to detect because the published literature does not always refer to the rôle an algebra system may have played in the work reported. The message to relativists would therefore seem to be to use whatever systems are locally available. It is only through increased usage that greater effort is likely to be expended in improving the existing systems. It is surely only a question of time before algebra systems become a commonplace tool for all those engaged in scientific research.

References

[1] Sources of Gravitational Radiation, ed. L. Smarr, Cambridge (1980).
[2] T. Piran, "Numerical Relativity", in: Recent Developments in General Relativity, ed. R. Ruffini (to appear).
[3] R. A. d'Inverno, "A Review of Algebraic Computing in General Relativity", in: General Relativity and Gravitation, volume 1, ed. A. Held, Plenum (1980).
[4] T. Piran, J. Comp. Phys. **35**, 254 (1980).
[5] J. M. Centrella, "Interacting Gravitational Shocks in Vacuum Plane-symmetric Cosmologies", in preparation 1979.
[6] J. M. Centrella, "Interacting Gravitational and Matter Shocks in Plane-symmetric Cosmologies", in preparation 1979.
[7] A. M. Anile, J. C. Miller and S. Motta, "Numerical Calculation of Formation and Damping of Fully Relativistic Strong Shocks". GR9 Abstracts, Volume 1 (1980).
[8] R. A. d'Inverno and J. Stachel, J. math. Phys. **19**, 2447 (1978).
[9] R. A. d'Inverno and J. Smallwood, "On a Covariant 2 + 2 Formulation of the Initial Value Problem in General Relativity". Phys. Rev. D (to appear).
[10] H. Müller zum Hagen and H. J. Seiffert, GRG **8**, 259 (1977).
[11] Private communication.
[12] G. V. Bicknell and R. N. Henriksen, Astrophys. J. **232**, 670 (1979).
[13] H. Friedrich, Ph. D. thesis, Hamburg University 1979.
[14] H. Friedrich, "On the Regular and Asymptotic Characteristic Initial Value Problem in General Relativity", GR 9 Abstracts, volume 1 (1980).
[15] Cheuk-Yin Wong, J. math. Phys. **12**, 70 (1971).
[16] P. A. Collins and R. M. Williams, Phys. Rev. **D 5**, 1908 (1972).
[17] P. A. Collins and R. M. Williams, Phys. Rev. **D 7**, 965 (1973); **10**, 3537 (1974).
[18] R. M. Williams and G. F. R. Ellis, "Regge Calculus and Observations", GR9 Abstracts, volume 1, (1980).
[19] J. E. Sammet, Commun. ACM. **9**, 555 (1966).
[20] A. C. Hearn, Commun. ACM. **14**, 511 (1971).
[21] D. C. Richardson, Ph. D. thesis, Bristol University 1966.
[22] R. H. Risch, Bull. Am. Math. Soc. **76**, 605 (1970).
[23] A. Krasiński and M. Perkowski, "ORTOCARTAN — A Computer Program for Calculating Curvature Tensors", GR9 Abstracts volume 1 (1980).
[24] D. W. Ebner, "GOEDEL — A Computer Language for Symbolic-Algebraic Calculations", GR9 Abstracts, volume 1 (1980).
[25] S. V. Tarasevich, "GRATOS — A FORTRAN-based System for General Relativity Symbolic Manipulations", GR 9 Abstracts, volume 1, (1980).
[26] H. I. Cohen, Ö. Leringe and Y. Sundblad, Gen. Rel. Grav. **7**, 269 (1976).
[27] R. A. d'Inverno, Comput. J. **12**, 124 (1969).
[28] R. A. d'Inverno, and R. A. Russell-Clark, Comput. J. **17**, 229 (1971).
[29] B. K. Harrison, Phys. Rev. **116**, 1285 (1959).
[30] R. A. d'Inverno and R. A. Russell-Clark, J. math. Phys. **12**, 1258 (1971).
[31] R. A. d'Inverno, J. math. Phys. **16**, 670 (1975).
[32] R. A. d'Inverno, J. math. Phys. **16**, 674 (1975).

[33] I. Hauser, Phys. Rev. Letters **33**, 1112 (1974).
[34] I. Robinson, Gen. Rel. Grav. **6**, 423 (1974).
[35] A. Held, Lett. Nuovo Cimento **11**, 545 (1974).
[36] R. A. d'Inverno and J. Smallwood, Gen. Rel. Grav. **9**, 195 (1978).
[37] R. A. d'Inverno and J. Smallwood, Gen. Rel. Grav. **9**, 215 (1978).
[38] I. Frick, "The Computer Algebra System SHEEP. What it can and cannot do in General Relativity", Stockholm University, Sweden 1977.
[39] R. A. d'Inverno and I. Frick, "Interacting with SHEEP", Gen. Rel. Grav. (to appear).
[40] Private communication.
[41] P. S. Florides, in "General Relativity: Papers in Honour of J. L. Synge", Oxford, 167 (1977).
[42] A. Karlhede, U. Lindström, J. A. Åman, "How to avoid Black Holes". Stockholm University, Sweden 1980.
[43] C. H., Brans, J. math. Phys. **6**, 94 (1965).
[44] A. Karlhede, "On a Coordinate Invariant Description of Riemannian Manifolds" (submitted to Gen. Rel. Grav.)
[45] A. Karlhede and J. Åman, "Classifying Geometries in General Relativity", GR 9 Abstracts, volume 1, (1980).
[46] W. Kinnersley, J. math. Phys. **10**, 1195 (1965).
[47] L. Hörnfeldt, "Indicial Tensor and Spinor Calculus with Computer", GR 9 Abstracts, volume 1, (1980).
[48] L. Hörnfeldt and R. Pavelle, "Gauss-Bonnet Invariants and Indicial Tensor Manipulation on MACSYMA", GR 9 Abstracts, volume 1, (1980).
[49] MACSYMA Reference Manual. Laboratory for Computer Science, M. I. T.
[50] R. Pavelle, "Applications of MACSYMA to problems in Gravitation and Differential Geometry". Proceedings of 1979 MACSYMA Users Conference, Washington.
[51] R. Pavelle, Gen. Rel. Grav. **7**, 383 (1976).
[52] J. P. Fitch, CAMAL User's Manual. Computer Laboratory. University of Cambridge 1975.
[53] J. Wainwright, "CAMAL Programs for GRT: A Users Guide", University of Waterloo, Canada 1977.
[54] J. P. Fitch, Ph. D. thesis, Cambridge University 1971.
[55] G. W. Gibbons and R. A. Russell-Clark, Phys. Rev. Letters **30**, 398 (1973).
[56] J. Wainwright, J. math. Phys. **18**, 672 (1977).
[57] D. A. Szafron and J. Wainwright, J. math. Phys. **18**, 1668 (1977).
[58] J. Wainwright, Gen. Rel. Grav. **8**, 799 (1977).
[59] S. J. Campbell and J. Wainwright, Gen. Rel. Grav. **8**, 987 (1977).
[60] Private communication.
[61] D. Kramer, H. Stephani, M. MacCallum, E. Herlt: Exact Solutions of Einstein's Field Equations. VEB Deutscher Verlag der Wissenschaften Berlin, 1980.
[62] S. T. C. Siklos, "The Equivalence Problem in General Relativity", Queen Mary College Preprint, (1980).

Address: Prof. Dr. R. A. d'Inverno
Fac. Math. Studies
Univ. of Southampton
Southampton S09 5NH
England

The Initial Value Problem and the Dynamics of Gravitational Fields[1])

J. E. Marsden (Berkeley)

This lecture will survey some of the recent advances that have been made in the dynamics of general relativity and other classical relativistic field theories. In addition, we shall indicate a few open problems that appear to be of basic interest.

1. Existence and Uniqueness Theorems for Geometrodynamics

The basic existence and uniqueness theorem states that Cauchy data on a spacelike hypersurface determines uniquely (up to spacetime diffeomorphisms) a piece of spacetime (filled with whatever matter or other fields are under consideration) containing the hypersurface. Moreover, it makes sense to look at the maximal development of such Cauchy data, just as it makes sense to look at maximal integral curves of ordinary differential equations.

The rigorous theory developing results of this type begins with Choquet-Bruhat [1948], [1952]. The subject as it existed up until about 1972 is adequately presented in Hawking and Ellis [1973]. Some of the key developments since then are as follows, in more or less chronological order:

(a) Fischer and Marsden [1972a] show how to write the evolution equations as a first-order symmetric hyperbolic system.

(b) Müller-zum-Hagen and Seifert [1977] study the characteristic initial value problem.

(c) Hughes, Kato and Marsden [1977] prove a conjecture of Hawking and Ellis, showing that the equations are well posed, with the metric in H^s, $s > 2.5$. (See also Fischer and Marsden [1979a].)

(d) For asymptotically flat spacetimes, Choquet-Bruhat and Christodoulou [1980] and Christodoulou [1980] prove well-posedness in the weighted Sobolev spaces of Nirenberg-Walker and Cantor ("SNWC spaces"; see Cantor [1979]). The crucial point here is to allow Hilbert spaces (cf. McOwen [1979]).

(e) Christodoulou and O'Murchadha [1980] solve the boost problem in SNWC spaces; i.e. they show that the piece of spacetime generated by the initial data is large enough at spatial infinity to include boosts. The methods may allow also for capturing a piece of Ω.

[1]) Research partially supported by the National Science Foundation.

Some open problems for 1.

(i) *Gauge Problem:* Current existence theory is based on the harmonic gauge of Choquet-Bruhat. Is there a direct proof valid in any gauge?

(ii) *Global Problem:* Find a gauge (such as the constant mean curvature gauge, i.e. $\varkappa$-gauge) in which global existence holds. Recent work of Christodoulou, Eardley and Moncrief on Yang-Mills fields and Maxwell Klein-Gordon fields gives one hope that the gravitational problem may be solvable. (See Segal [1979], Moncrief [1980a, b, c] and Eardley and Moncrief [1980]). For globally hyperbolic spacetimes with a compact Cauchy surface (cosmological case) global existence in a constant mean curvature gauge implies that the evolution in that gauge captures the entire spacetime; see below and Marsden and Tipler [1980]. For non-compact Cauchy surfaces, this need not be true; see Eardley and Smarr [1979].

(iii) *Boundaries and Gravitational Shocks.* Try to lower s even below 2.5 in the Cauchy problem to allow for jump discontinuities in the second derivatives of the metric ($s = 2.5$ is the crucial value, below which jumps are allowed). This would allow gravitational shocks. The solutions are presumably non-unique if $s < 2.5$ and the physically correct ones are picked out by some kind of entropy condition, as one does in gas dynamics. Can recent advances in geometric optics and Fourier integral operators (Guillemin and Sternberg [1977]) be used in the study of gravitational waves and shocks?

2. Hamiltonian Structures

The Hamiltonian formalism in general relativity goes back to Choquet-Bruhat, Dirac, Bergmann and Arnowitt-Deser and Misner. This formalism (referred to commonly as the ADM formalism) is found in, for example, Misner, Thorne and Wheeler [1973].

This formalism can be exhibited as follows (Fischer-Marsden [1976, 1979]). Let a slicing of spacetime (V, ${}^{(4)}g$) be given, based on a 3-manifold M. This slicing determines a curve $g(\lambda)$ of Riemannian metrics on M and a curve of symmetric tensor densities $\pi(\lambda)$ (the conjugate momentum). Let the slicing have a lapse function N and a shift vector field X. Einstein's vacuum equations Ein $({}^{(4)}g) = 0$ (the Einstein tensor formed from ${}^{(4)}g$) are then equivalent to the evolution equations in adjoint form

$$\frac{\partial}{\partial\lambda}\begin{pmatrix} g \\ \pi \end{pmatrix} = \mathbb{J} \circ D\Phi(g, \pi)^* \begin{pmatrix} X \\ N \end{pmatrix}; \qquad \mathbb{J} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

together with the constraints

$$\Phi(g, \pi) = 0$$

where $\Phi(g, \pi) = \big(\mathscr{H}(g, \pi), \mathscr{J}(g, \pi)\big)$ is the super energy-momentum.

The choice of slicing is a gauge choice, and one may wish to determine it along with the dynamics. In particular, the constant mean curvature gauge is especially interesting, as we have already noted. As was indicated in Professor Wheeler's lecture and in Qadir and Wheeler [1980], this gauge has the property that its spacelike hypersurfaces tend to avoid singularities. If one can show that the mean curvature $\varkappa$ runs from $-\infty$ to ∞ and M is compact (closed universes) then the foliation fills out the whole Cauchy development and in fact this development is a "Wheeler universe" (see Tipler and Marsden [1980]).

Some other developments of interest in the Hamiltonian formalism are:

(a) The Hamiltonian and symplectic structures are investigated directly from the four dimensional point of view in Kijowski and Szczyrba [1976], and Kijowski and Tulczyjew [1979].

(b) There has been development of the idea that the constraint $\Phi = 0$ is the same as the vanishing of the Noether current generated by the gauge group of relativity i. e. all diffeomorphisms of V (equalling the identity at infinity for open universes); for relativity, see Fischer and Marsden [1972b], for gauge theories, see Cordero and Teitelboim [1976], Moncrief [1977] and Arms [1978].

(c) The Poincaré group at infinity or the BMS group have Noether currents of interest as well, (although we do not set them zero) such as the ADM energy-momentum tensor or the BMS energy-momentum tensor; see Regge-Teitelboim [1974] and Ashtekar and Streuble [1980].

(d) How $\varkappa$-slicings fit together with the BMS group and gravitational radiation has been investigated by Stumbles [1980]. (For related information on $\varkappa$-slicings, see Choquet-Bruhat, Fischer and Marsden [1979], Eardley and Smarr [1979], Marsden and Tipler [1980] and Treibergs [1980] and references therein).

(e) Teitelboim [1977] and Pilati [1978] have investigated the geometrodynamics of supergravity. Bao [1981] has put it onto the adjoint form above.

Some open problems for 2.

(i) Find sufficient conditions on a relativistic field theory with a given gauge group to ensure that the constraints in a Dirac analysis will be the zero level of the corresponding Noether current. (This is true for all the examples mentioned above).

(ii) How is the classical Noether constraint "color charge = 0" for Yang-Mills fields on Minkowski space related to quark confinement? (See Arms, Marsden and Moncrief [1980] for some discussion).

(iii) Is it true that the long time dynamics for a typical relativistic field theory is chaotic? Is the Kolmogorov-Arnold-Moser theory relevant? (Recently, Barrow has embarked on a very enlightening investigation of Misner's mixmaster model from the point of view of chaotic dynamical systems).

3. Spaces of Solutions and Linearization Stability

Let V be a fixed four manifold and let $\mathscr{E}$ be the set of all globally hyperbolic Lorentz metrics $g = {}^{(4)}g$ that satisfy the vacuum Einstein equations $\operatorname{Ein}(g) = 0$ on V (plus some additional technical smoothness conditions). Let $g_0 \in \mathscr{E}$ be a given solution. We ask: what is the structure of $\mathscr{E}$ in the neighborhood of g_0?

There are two basic reasons why this question is asked. First of all, it is relevant to the problem of finding solutions to the Einstein equations in the form of a perturbation series:

$$g(\lambda) = g_0 + \lambda h_1 + \frac{\lambda^2}{2} h_2 + \cdots$$

where λ is a small parameter. If $g(\lambda)$ is to solve $\operatorname{Ein}\big(g(\lambda)\big) = 0$ identically in λ then clearly h_1 must satisfy the *linearized Einstein equations*:

$$\mathrm{D}\operatorname{Ein}(g) \cdot h_1 = 0$$

where $\mathrm{D}\operatorname{Ein}(g)$ is the derivative of the mapping $g \mapsto \operatorname{Ein}(g)$. For such a perturbation series to be possible, is it sufficient that h_1 satisfy the linearized Einstein equations, i.e. is h_1 necessarily a direction of *linearized stability*? We shall see that in general the answer is no, unless drastic additional conditions hold. The second reason why the structure of $\mathscr{E}$ is of interest is in the problem of quantization of the Einstein equations. Whether one quantizes by means of direct phase space techniques (due to Dirac, Segal, Souriau and Kostant in various forms) or by Feynman path integrals, there will be difficulties near places where the space of classical solutions is such that the linearized theory is *not* a good approximation to the nonlinear theory.

The dynamical formulation mentioned in § 2 is crucial to the analysis of this problem. Indeed, the essence of the problem reduces to the study of structure of the space of solutions of the constraint equations $\Phi(g, \pi) = 0$.

As we shall see, the answer to these questions is this: $\mathscr{E}$ has a conical or quadratic singularity at g_0 if and only if there is a nontrivial Killing field for g_0 that belongs to the gauge group generating $\Phi = 0$ (thus, the flat metric on $T^3 \times \mathbb{R}$ has such Killing fields, but the Minkowski metric has none.) When $\mathscr{E}$ has such a singularity, we speak of a bifurcation in the space of solutions.

(a) Brill and Deser [1973] considered perturbations of the flat metric on $T^3 \times \mathbb{R}$ and discovered the first example of trouble in perturbation theory. They found, by going to a second order perturbation analysis, that they had to readjust the first order perturbations in order to avoid inconsistencies at second order. This was the first hint of a conical structure for $\mathscr{E}$ near solutions with symmetry.

(b) Fischer and Marsden [1973] found general sufficient conditions for $\mathscr{E}$ to be a manifold in terms of the Cauchy data for vacuum spacetimes.

(c) Choquet-Bruhat and Deser [1973] proved a version of the theorem that $\mathscr{E}$ is a manifold near Minkowski space, which was later improved by Choquet-Bruhat, Fischer and Marsden [1979].

(d) Moncrief [1975a] showed that the sufficient conditions derived by Fischer and Marsden for the compact case where equivalent to the requirement that (V, g_0) have no Killing fields. This then led to the link between symmetries and bifurcations.

(e) Moncrief [1975b] discovered the general splitting of gravitational perturbations generalizing Deser's [1967] decomposition. The further generalization to momentum maps (general Noether currents) was found by Arms, Fischer and Marsden [1975]. This then applies to other examples such as gauge theory and also gives York's decomposition (York [1974]) as special cases.

(f) D'Eath [1976] obtained the basic linearization stability results for Robertson-Walker universes.

(g) Moncrief [1976] discovered the spacetime significance of the second order conditions that arise when one has a Killing field and identified them with conserved quantities of Taub [1970]. Arms and Marsden [1979] showed that the second order conditions for compact spacelike hypersurfaces are nontrivial conditions.

(h) The description of the conical singularity in $\mathscr{E}$ near a spacetime with symmetries is due to Fischer, Marsden and Moncrief [1980] for one Killing field and to Arms, Fischer, Marsden and Moncrief [1981] in the general case.

(i) Moncrief [1977], Coll [1975] and Arms [1977, 1979] obtained the basic results for pure gauge theories and electromagnetism and gauge theories coupled to gravity.

(j) An abstract theory valid for arbitrary momentum maps was developed by Arms, Marsden and Moncrief [1980].

(k) Moncrief [1978] investigated the quantum analogues of linearization stabilities. Using $T^3 \times \mathbb{R}$, he shows that, unless such conditions are imposed, the correspondence principle is violated.

For vacuum gravity, let us state one of the main results in the cosmological case: suppose g_0 has a *compact* spacelike hypersurface $M \subset V$. (Actually we require the existence of at least one of constant mean curvature for technical reasons). Let S_{g_0} be the Lie group of isometries of g_0 and let k be its dimension.

Theorem.

1. $k = 0$, then $\mathscr{E}$ is a smooth manifold in a neighborhood of g_0 with tangent space at g_0 given by the solutions of the linearized Einstein equations.
2. If $k > 0$ then $\mathscr{E}$ is *not* a smooth manifold at g_0. A solution h_1 of the linearized equations is tangent to a curve in $\mathscr{E}$ if and only if h_1 is such that Taub conserved quantities vanish; i.e. for every Killing field X for g_0,

$$\int_M X \cdot [\mathrm{D}^2 \operatorname{Ein}(g_0) \cdot (h_1, h_1)] \cdot Z \, d\mu_M = 0$$

where Z is the unit normal to the hypersurface M and "." denotes contraction with respect to the metric g_0.

All explicitly known solutions possess symmetries, so while 1. is "generic", 2. is what occurs in examples. This theorem gives a complete answer to the perturbation question: a perturbation series is possible if and only if all the Taub quantities vanish.

Let us give a brief abstract indication of why such second order conditions should come in. Suppose X and Y are Banach spaces and $F: X \to Y$ is a smooth map. Suppose $F(X_0) = 0$ and $x(\lambda)$ is a curve with $x(0) = x_0$ and $F(x(\lambda)) \equiv 0$. Let $h_1 = x'(0)$ so by the chain rule $DF(x_0) \cdot h_1 = 0$. Now suppose $DF(x_0)$ is not surjective and in fact suppose there is a linear functional $l \in Y^*$ orthogonal to its range: $\langle l, DF(x_0) \cdot u\rangle = 0$ for all $u \in X$. By differentiating $F(x(\lambda)) = 0$ twice at $\lambda = 0$, we get

$$D^2F(x_0) \cdot (h_1, h_1) + DF(x_0) \cdot x''(0) = 0.$$

Applying l gives

$$\langle l, D^2F(x_0) \cdot (h_1, h_1)\rangle = 0$$

which are necessary second order conditions that must be satisfied by h_1.

It is by this general method that one arrives at the Taub conditions. The issue of whether or not these conditions are sufficient is much deeper requiring extensive analysis and bifurcation theory (for $k = 1$ the Morse lemma is used, while for $k > 1$ the Kuranishi deformation theory is needed see Kuranishi [1965], Atiyah, Hitchin and Singer [1978] and § 4 below).

Some open problems for 3.

(i) Is the above phenomenon a peculiarity about vacuum gravity or is there an abstract theorem applicable to a broad class of relativistic field theories? The examples which have been and are being worked out suggest that the latter is the case. Good examples are the Yang-Mills equations for gauge theory (Moncrief [1977], Arms [1979]) the Einstein-Dirac equations (cf. Nelson and Teitelboim [1978]), the Einstein-Euler equations (Bao and Marsden [1981]) and supergravity (Pilati [1978], Bao [1981]). In each of these examples there is a gauge group playing the role of the diffeomorphism group of spacetime for vacuum gravity. This gauge group acts on the fields; when it fixes a field, it is a *symmetry* for that field. The relationship between symmetries of a field and singularities in the space of solutions of the classical equations is then as it is for vacuum gravity.

For this program to carry through, one first writes the four dimensional equations as Hamiltonian evolution equations plus constraint equations by means of the 3 + 1 procedures of Dirac. The constraint equations then must

1. *be the Noether conserved quantities for the gauge group and* 2. *satisfy some technical ellipticity conditions:* $(D\Phi)^*$ *must be an elliptic operator*. As is already been mentioned, for 1, it may be necessary to shrink the gauge group somewhat, especially for spacetimes that are not spatially compact. For example the isometries of Minkowski space do not belong to the gauge group generating the constraints but rather they generate the total energy-momentum vector of the

spacetime ... that this vector is time-like is the now famous positive energy problem ... see Brill and Deser [1978], Choquet/Bruhat, Fischer and Marsden [1979], Deser and Teitelboim [1976] and Schoen and Yau [1979, 1980].

(ii) In the space of solutions, the kernel of the symplectic form coincides with the infinitesimal gauge transformations (this follows from Moncrief's decomposition). Therefore, one can construct the space of true degrees of freedom, the quotient of $\mathscr{E}$ by the gauge group. Using Marsden-Weinstein [1974], one proves that this quotient is a smooth symplectic manifold near points where $\mathscr{E}$ is smooth. This leaves open the question: what is $\mathscr{E}$/(gauge group) like near points of symmetry, where $\mathscr{E}$ is singular?

(iii) How should one treat the Schwarzschild solution in the context of linearization stability? Do singularities in the space of solutions affect spacetime singularities in the sense of Hawking and Penrose? Do they affect Cauchy horizons?

4. Bifurcations of Momentum Maps

The role of the constraint equations as the zero set of the Noether conserved quantity of the gauge group leads one to investigate zero sets of the conserved quantities associated with symmetry groups rather generally. One goal is to begin answering question (i) in the previous section. This topic is of interest not only in relativistic field theories, but in classical mechanics too. For example the set of points in the phase space for n particles in $\mathbb{R}^3$ corresponding to zero total angular momentum is an interesting and complicated set, even for $n = 2$!

We shall present just a hint of the relationship between singularities and symmetries. The full story is a long one; one finally ends up with an answer similar to that in vacuum relativity. We refer to Arms, Marsden and Moncrief [1980] for additional details.

First we need a bit of notation (see Abraham and Marsden [1978], Chapter 4). Let M be a manifold and let a Lie group G act on M. Associated to each element ξ in the Lie algebra $\mathfrak{G}$ of G, we have a vector field ξ_M naturally induced on M. We shall denote the action by $\Phi : G \times M \to M$ and we shall write $\Phi_g : M \to M$ for the transformation of M associated with the group element $g \in G$. This $\xi_M(x) = \frac{d}{dt} \Phi_{\exp(t\xi)}(x)|_{t=0}$.

Now let (P, ω) be a symplectic manifold, so ω is a closed (weakly) non-degenerate two-form on P and let Φ be an action of a Lie group G on P. Assume the action is symplectic: i.e. $\Phi_g^*\omega = \omega$ for all $g \in G$. A *momentum mapping* is a smooth mapping $J : P \to \mathfrak{G}^*$ such that

$$\langle dJ(x) \cdot v_x, \xi \rangle = \omega\big(\xi_P(x), v_x\big)$$

for all $\xi \in \mathfrak{G}$, $v_x \in T_xP$ where $dJ(x)$ is the derivative of J at x, regarded as a linear map of T_xP to $\mathfrak{G}^*$ and $\langle,\rangle$ is the natural pairing between $\mathfrak{G}$ and $\mathfrak{G}^*$.

A momentum map is *Ad*-equivariant* when the following diagram commutes for each $g \in G$:

$$\begin{array}{ccc} P & \xrightarrow{\Phi_g} & P \\ \Big\downarrow J & & \Big\downarrow J \\ \mathfrak{G}^* & \xrightarrow[\mathrm{Ad}^*_{g-1}]{} & \mathfrak{G}^* \end{array}$$

where Ad^*_{g-1} denotes the co-adjoint action of G on $\mathfrak{G}^*$. If J is Ad* equivariant, we call (P, ω, G, J) a *Hamiltonian* G-*space.*

Momentum maps represent the (Noether) conserved quantities associated with symmetry groups acting on phase space. This topic is of course a very old one, but it is only with more recent work of Souriau and Kostant that a deeper understanding has been achieved.

See Fischer and Marsden [1979] for the sense in which the map Φ described in § 2 is the momentum map associated with the group of diffeomorphisms of spacetime. See Moncrief [1977] and Arms [1979] for the corresponding result for gauge theory.

Let $S_{x_0} =$ (the component of the identity of) $\{g \in G \mid gx_0 = x_0\}$, called the symmetry group of x_0. Its Lie algebra is denoted $\mathfrak{s}_x$, so

$$\mathfrak{s}_{x_0} = \{\xi \in \mathfrak{G} \mid \xi_P(x_0) = 0\}.$$

Let (P, ω, G, J) be a Hamiltonian G-space. If $x_0 \in P$, $\mu_0 = J(x_0)$ and if

$$dJ(x_0) : T_x P \to \mathfrak{G}^*$$

is surjective (with split kernel), then locally $J^{-1}(\mu_0)$ is a manifold and $\{J^{-1}(\mu) \mid \mu \in \mathfrak{G}^*\}$ forms a regular local foliation of a neighbourhood of x_0. Thus, when $dJ(x_0)$ fails to be surjective, the set of solutions of $J(x) = 0$ could fail to be a manifold.

Theorem. $dJ(x_0)$ is surjective if and only if $\dim S_{x_0} = 0$; i.e. $\mathfrak{s}_{x_0} = \{0\}$.

Proof. $dJ(x_0)$ fails to be surjective if there is a $\xi \neq 0$ such that $\langle dJ(x_0) \cdot v_{x_0}, \xi \rangle = 0$ for all $v_{x_0} \in T_{x_0}P$. From the definition of momentum map, this is equivalent to $\omega_{x_0}(\xi_P(x_0), v_{x_0}) = 0$ for all v_{x_0}. Since ω_{x_0} is non-degenerate, this is, in turn equivalent to $\xi_P(x_0) = 0$; i.e. $\mathfrak{s}_{x_0} \neq \{0\}$.

One then goes on to study the structure of $J^{-1}(\mu_0)$ when x_0 *has* symmetries, by investigating second order conditions and using methods of bifurcation theory. It turns out that, as in relativistic field theories, $J^{-1}(\mu_0)$ has quadratic singularities characterized by the vanishing of second order conditions. The connection is not an accident since the structure of the space of solutions of a relativistic field theory is determined by the vanishing of the momentum map associated with the gauge group of that theory.

Some open problems for 4.

(i) All the results obtained so far on spaces of solutions are local. What is the global structure? Is there a global Morsetype theory for momentum maps?

(ii) Much current work on Yang-Mills fields and the twistor program for gravity utilize a Euclideanized viewpoint. Some routine calculations show that in such a context the connection between symmetries and bifurcations is lost. (In particular, the symmetries discussed by Rebbi and Jackiw [1976] are *not* related to Euclidean linearization instabilities.) What has become of the difficulties with perturbation series and quantization encountered in the Lorentz context?

(iii) Bifurcation theory exploits connections between symmetry and bifurcation to study phenomena like pattern formation. See for example, Golubitsky and Schaeffer [1979] and Sattinger [1980]. Can one use this theory in relativity to study physical consequences of breaking the symmetry of a solution of a relativistic field theory?

References

R. Abraham and J. Marsden [1978]. Foundations of Mechanics. Second Edition. Addison-Wesley.

J. Arms [1977]. Linearization stability of the Einstein-Maxwell system, J. math. Phys., 18, 830–3.

— [1979]. Linearization stability of gravitational and gauge fields, J. math. Phys. **20**, 443–453.

— [1980]. The structure of the solution set for the Yang-Mills equations (preprint).

J. Arms, A. Fisher and J. Marsden [1975]. Une approche symplectique pour des théorèmes de decomposition en geométrie ou relativité générale, C.R. Acad. Sci. Paris, **281**, 517–520.

J. Arms, A. Fisher, J. Marsden and V. Moncrief [1980]. The structure of the space of solutions of Einstein's equations; II Many Killing fields (in preparation).

J. Arms and J. Marsden [1979]. The absence of Killing fields is necessary for linearization stability of Einstein's equations, Ind. Univ. Math. J. **28**, 119–125.

J. Arms, J. Marsden and V. Moncrief [1980]. Bifurcations of momentum mappings Commun. math. Phys. (to appear).

A. Ashtekar and M. Streubel [1980]. Symplectic geometry of radiative modes and conserved quantities at null infinity (preprint).

M. F. Atiyah, M. J. Hitchin and I. M. Singer [1978]. Self-duality in four-dimensional Riemannian geometry. Proc. royal Soc. London, **A 362**, 425–461.

D. Bao [1981]. Thesis, Berkeley.

D. Bao and J. Marsden [1981]. The dynamics of relativistic continuum mechanics (in preparation).

D. Brill and S. Deser [1968]. Variational methods and positive energy in relativity, Ann. Phys. **50**, 548–570.

M. Cantor [1979]. Some problems of global analysis on asymptotically simple manifolds. Comp. Math. **38**, 3–35.

— [1980]. Elliptic operators and the decomposition of tensor fields (preprint).

Y. Choquet (Foures)-Bruhat [1948]. Sur l'intégration des équations d'Einstein. C. R. Acad. Sci Paris **226**, 1071. (See also J. Ration. Mech. Anal., **5** (1956) 951–96.)

— [1952]. Théorème d'existence pour certains systems d'equations aux derivees partielles non lineaires, Acta math. **88**, 141–225.

Y. Choquet-Bruhat and D. Christodoulou [1980]. Elliptic systems in Hilbert spaces on manifolds which are Euclidean at infinity (preprint).

Y. Choquet-Bruhat and S. Deser [1972]. Stabilité initiale de l'espace temps de Minkowski, C. R. Acad. Sci. Paris **275**, 1019—1927.

— [1973]. On the stability of flat space, Ann. Phys. **81**, 165—168.

Y. Choquet-Bruhat, A. Fisher and J. Marsden [1979]. Maximal Hypersurfaces and Positivity of Mass, in Isolated Gravitating Systems and General Relativity, J. Ehlers, ed., Italian Physical Society, 322/395.

D. Christodoulou [1980]. The boost problem for weakly coupled quasilinear hyperbolic systems of the second order (preprint).

— D. Christodoulu and N. O'Murchadha [1980]. The boost problem in general relativity (preprint).

B. Coll [1975). Sur la détermination, par des donnés de Cauchy, des champs de Killing admis par un espace-temps, d'Einstein-Maxwell. C. R. Acad. Sci. Paris **281**, 1109—12.

— [1976]. Sur la stabilité linéaire des equations d'Eintein du vide. C. R. Acad. Sci. Paris **282**, 247—250.

— [1978]. On the evolution equations for Killing fields. J. math. Phys., **18**, 1918—22.

P. Cordero and C. Teitelboim [1976]. Hamiltonian treatment of the spherically symmetric Einstein-Yang-Mills system. Ann. Phys. (N.Y.), **100**, 607—31.

P. D. D'Eath [1976]. On the existence of perturbed Robertson-Walker universes. Ann. Phys. (N.Y.), **98**, 237—63.

S. Deser [1967]. Covariant decomposition of symmetric tensors and the gravitational Cauchy problem, Ann. Inst. H. Poincaré VII, 149—188.

S. Deser and C. Teitelboim [1976]. Supergravity has positive energy. Phys. Rev. Letters **39**, 249—52.

D. Eardley and V. Moncrief [1980]. Global existence for the Yang-Mills equations.

D. Eardley and L. Smarr [1979]. Time functions in general relativity I, marginal bound collapse, Phys. Rev. **D 19**, 2239—2245,

A. Fischer and J. Marsden [1972a]. The Einstein evolution equations as a first-order symmetric hyperbolic quasilinear system. Commun. math. Phys. **28**, 1—38.

— [1972b]. The Einstein equations of evolution — a geometric approach. J. math. Phys., **13**, 546—68.

— [1973]. Linearization stability of the Einstein equations. Bull. Amer. Math. Soc., **79**, 995—1001.

— [1976]. A new Hamiltonian structure for the dynamics of general relativity, J. Grav. Gen. Rel. **7**, 915—920.

— [1979a]. The initial value problem and the dynamical formulation of general relativity, in "General Relativity, An Einstein Centenary Survey", ed. S. Hawking and W. Israel, Cambridge, Ch. 4.

— [1979b]. Topics in the dynamics of general relativity, in "Isolated graviting systems in General Relativity", J. Ehlers, ed., Italian Physical Society, 322—395.

M. Golubitsky, and D. Schaeffer [1979]. Imperfect bifurcation in the presence of symmetry, Commun. math. Phys. **67**, 205—232.

V. Guillemin and S. Steinberg [1977]. Geometric asymptotics. Amer. Math. Soc. Survey, vol. 14, American Math. Society, Providence, R.I.

S. Hawking and G. Ellis [1973]. The Large Scale Structure of Spacetime, Cambridge University Press.

T. Hughes, T. Kato and J. Marsden [1977]. Well-posed quasi-linear hyperbolic systems with applications to nonlinear elastodynamics and general relativity, Arch. Rat. Mech. An. **63**, 273—294.

J. Kijowski and W. Szczyrba [1976]. A canonical structure for classical field theories. Commun. math. Phys. **46**, 183—205.

J. Kijowski and W. Tulczyjew [1979]. A symplectic framework for field theories, Springer Lecture Notes in Physics 107.

M. Kuranishi [1975]. New proof for the existence of locally complete families of complex structures. Proc. of Conf. on Complex Analysis, (A. Aeppli et al., eds), Springer.

J. Marsden and F. Tipler [1980]. Maximal hypersurfaces and foliations of constant mean curvature in general relativity. Physics Reports (to appear).

J. Marsden and A. Weinstein [1974]. Reduction to symplectic manifolds with symmetry, Rep. Math. Phys. **5**, 121—130.

R. C. McOwen [1979]. The behavior of the Laplacian on weighted Sobolev spaces, Comm. pure appl. Math. **32**, 781—793.

C. Misner, K. Thorne and J. Wheeler [1973]. Gravitation. Freeman, San Francisco.

V. Moncrief [1975a]. Spacetime symmetries and linearization stability of the Einstein equations, J. math. Phys. **16**, 493—498.

— [1975b]. Decompositions of gravitational perturbations, J. math. Phys. **16**, 1556—1560.

— [1976]. Spacetime symmetries and linearization stability of the Einstein equations II J. math. Phys. **17**, 1893—1902.

— [1977]. Gauge symmetries of Yang-Mills fields, Ann. Phys. **108**, 387—400.

— [1978]. Invariant states and quantized gravitational perturbations, Phys. Rev. **D 18**, 983—989.

— [1979]. Quantum linearization instabilities, J. Gen. Rel. & Grav. **10**, 93—97.

— [1980a]. Global existence of Maxwell-Klein-Gordon fields in $(2 + 1)$-dimensional spacetime. J. math. Phys. **21**, 2291—2296.

— [1980b]. Global properties of Gowdy spacetimes with $T^3 \times \mathbb{R}$ topology (preprint).

— [1980c]. An infinite dimensional family of vacuum cosmological models with TAUB-NUT type extensions (preprint).

H. Müller zum Hagen and H. J. Seifert [1977]. On the characteristic initial value and mixed problems, GRG 8, 259—301.

J. Nelson and C. Teitelboim [1977]. Hamiltonian for the Einstein-Dirac field, Phys. Letters **69B**, 81—84, and Ann. Phys. (1977).

L. Nirenberg and H. F. Walker. The Null Spaces of Elliptic Partial Differential Operators in $\mathbb{R}^n$, J. Math. An. & Appl. **42** (1973) 271—301.

N. O'Murchadha and J. York [1973]. Existence and uniqueness of solution of the Hamiltonian constraint of general relativity on compact manifolds. J. math. Phys. **14**, 1551—1557.

M. Pilati [1978]. The canonial formulation of supergravity, Nucl. Physics **B 132**, 138—154.

A. Qadir and J. A. Wheeler [1980]. Unification of the black hole with the big crunch (preprint).

P. Rebbi and J. Jackiw [1976]. Phys. Rev. **D 14**, 517—525.

T. Regge and C. Teitelboim [1974]. Role of surface integrals in the Hamiltonian formulation of general relativity. Ann. Phys. (N.Y.) **88**, 286—318.

D. H. Sattinger [1979]. Group theoretic methods in bifurcation theory, Springer Lecture Notes in Math. //762.

R. Schoen and S.-T. Yau [1979]. On the proof of the positive mass conjecture, Commun. math. Phys. **65**, 45—76.

— [1980]. On the proof of the positive mass conjecture, II (preprint).

I. Segal [1979]. The Cauchy problem for the Yang-Mills equations, J. Funct. An. **33**, 175—194.

S. Stumbles [1980]. Hypersurfaces of constant mean extrinsic curvature (preprint).

W. Szczyryba [1976]. A symplectic structure of the set of Einstein metrics: a canonical formalism for general relativity, Commun. math. Phys. **51**, 163—182.

A. Taub [1980]. Variational Principles in General Relativity, C.I.M.E. Bressanone, 206—300.

S. Teitelboim [1977]. Surface integrals as symmetry generators in supergravity theory, Phys. Letters **69B**, 240—244.

A. E. Treibergs [1980]. Entire spacelike hypersurfaces of constant mean curvature in Minkowski space. Thesis, Stanford.

J. W. York [1971]. Gravitational degrees of freedom and the initial-value problem. Phys. Rev. Letters 26, 1656—8.

— [1974]. Covariant decompositions of symmetric tensors in the theory of gravitation. Ann. Inst. H. Poincaré 21, 319—32.

Address: Prof. Dr. J. E. Marsden
Department of Mathematics
University of California
Berkeley, California 94720
USA

The Present State of the Twistor Programme

R. S. Ward (Dublin)

1. Introduction

The twistor programme contains several different aspects. To begin with, it has the general aim of providing an alternative framework for relativistic physics, one in which the underlying space is not space-time, but twistor space. According to this point of view, space-time points are derived rather than primary objects, and the possibility is opened up of formulating a theory of quantum gravity in which space-time points are "fuzzy", i.e. subject to quantum-mechanical uncertainties [1, 2]. The first step in this programme is that of translating standard space-time physics into the language of twistor theory. The twistor programme is still at this "translation" stage, although the new language has already provided several indications as to how one might modify (and perhaps improve) the conventional theory.

One of the most remarkable features that have so far appeared is the way in which certain classical field equations become considerably simpler when translated into the twistor picture. This simplification is closely tied up with the fact that twistor space is a *complex* manifold, and emphasizes the importance of complex numbers and holomorphic (i.e. complex-analytic) functions in the theory. Most of my lecture will be concerned with this aspect of the twistor programme, but I should at least mention two other areas of current research: elementary particle theory [3, 4], and the translation of Feynman diagrams into "twistor diagrams" [2, 4, 5].

2. Flat-space Twistor Theory

Let us begin by recalling the standard construction of the twistor space corresponding to Minkowski space-time M. More detailed descriptions may be found in [2, 6].

Let N denote the space of (unscaled) null geodesics in M. As a real manifold, N is diffeomorphic to $R^3 \times S^2$. A point x in M may be represented by the sphere's worth of null geodesics through x (i.e. by its null cone), so x is represented in N by a 2-sphere S_x (see fig. 1). The crucial step now is to imbed the five-real-dimensional space N in a three-complex-dimensional space T called (projective) twistor space. There is a completely natural, Poincaré-invariant way of doing this. One of the important consequences of having this structure is that it identifies those spheres in N which correspond to space-time points x. If we look for smooth spheres in N which belong to the appropriate homology class, then there are an infinite-parameter family of them; but only four real parameters' worth of these are holomorphic curves (i.e. one-dimensional complex-analytic submanifolds) in the complex manifold T, and this four-parameter family consists exactly of the spheres S_x which correspond to space-time points.

So, given the spaces N and T, one can "reconstruct" the space-time manifold M. Furthermore, the conformal structure (i.e. the null cones) of M automatically come out of the construction. For it is clear from the correspondence that

$$x \ \& \ y \text{ are null-separated} \Leftrightarrow S_x \ \& \ S_Y \text{ intersect.} \tag{1}$$

This is perhaps the most crucial and important feature of the space-time twistor space correspondence.

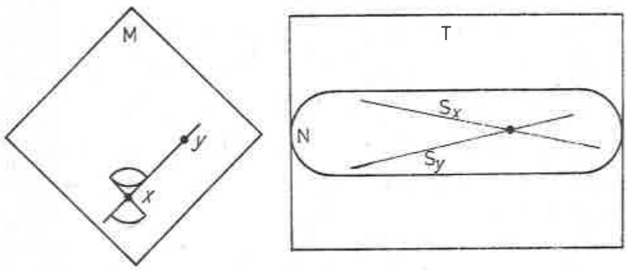


Fig. 1

Another striking consequence of the complex structure of T becomes apparent when we transform relativistic field equations from the space-time picture into the twistor-space picture. The simplest example is the wave equation $\Box\Phi = 0$. What one arrives at is the following.

Let f be a holomorphic function on some subset of T. Let f_x denote the restriction of f to the sphere S_x; thus f_x is a holomorphic function on some subset of the Riemann sphere (extended complex plane) S_x. Let γ be a closed contour in this subset; i.e. γ is a circle on S_x which avoids the places where f_x is not defined. Put

$$\varphi(x) = \oint_\gamma f_x. \tag{2}$$

Then $\varphi(x)$ is a solution of the wave equation, and *every* real-analytic solution of the wave equation arises in this way [2].

One may think about this result as follows. The general solution $\varphi(x)$ of $\Box\varphi = 0$ in four-dimensional space-time is in effect one or more *free* functions of *three* variables. For example, initial data on a spacelike hypersurface determines a solution throughout space-time. Similarly, the function f above is a free function of three variables, and determines a solution via the formula (2). (Admittedly, f is a function of three *complex* variables, but since f is holomorphic, it is correct to count these as three rather than six.)

Remarks

(i) We see, therefore, that the complex structure of twistor space, combined with the geometry of the space-time/twistor space correspondence, leads to a neat solution procedure for the wave equation. Much more remarkable is the fact that there are analogous solution procedures for certain *nonlinear* equations, as we shall see below.

(ii) The function f above is a rather awkward object to deal with, because it is not defined everywhere on T (it has singularities, in general) and because many different f's will give rise to the same $\varphi(x)$. It turns out that f should properly be regarded as an element of a *sheaf cohomology group* [4, 7].

(iii) The points of the space N correspond, of course, to null geodesics in space-time. But what about the points of T? They too can be interpreted in space-time: the most useful interpretation is to think of the points of T as corresponding to certain complex 2-planes, called α-planes, in *complexified* Minkowski space-time [2, 6].

(iv) The hypersurface N divides twistor space T into two halves, T^+ and T^- (see fig. 1). The holomorphic curves in T which lie entirely in T^+ correspond to the points of the *forward tube* [8] in complexified space-time [2]. As a consequence of this, the decription of fields $\varphi(x)$ which are of positive frequency, and the decomposition of fields into positive and negative frequencies, is particularly natural in the twistor picture.

3. Twistor solution of differential equations

When one transforms field equations from the space-time picture into the twistor-space picture, it sometimes turns out that a drastic simplification occurs: in fact, that the partial differential equation disappears altogether. We saw above that this happens to the scalar wave equation in flat space-time, and several more examples will be mentioned in this section. It seems that this phenomenon (in its simplest form) occurs only when the differential equations are of a rather simple type: namely when they satisfy *Huygens' Principle.*

Huygens' Principle may be stated as follows. If the field φ satisfies a hyperbolic field equation, then its value $\varphi(x)$ at some field point x may be expressed as a functional of the initial data on D_x, the intersection of the initial-data surface S and the causal past of the point x (see fig. 2). If the field equation is such that $\varphi(x)$ depends only on the initial data on the *boundary* ∂D_x of D_x, then we say that Huygens' Principle is satisfied. Qualitatively, it means that the field propagates "cleanly" along light rays, without back-scattering.

The standard example of a "Huygens" equation is the one we encountered in the previous section, namely the wave equation in flat space-time. The next examples that occur to mind are the equations for massless spinning free fields, such as Maxwell's equations. Here again one finds that the general solution of the equations in space-time corresponds naturally to "free" holomorphic data in twistor space [2, 4, 7].

So far, all the equations mentioned have been linear. There are also nonlinear equations which satisfy Huygens' Principle; the best-known example is the self-dual Yang-Mills equation, where one has a gauge field which is required to be self-dual (or anti-self-dual). In this case, it turns out that the self-dual gauge field in space-time corresponds to a holomorphic vector bundle over twistor space [9, 10]. Once again the partial differential equation in space-time disappears on going over to the twistor picture; and so (as a practical bonus) the twistor translation programme has presented us with a procedure for generating solutions of these nonlinear equations. This procedure has been successfully applied to the search for instanton [10—13] and monopole [14] solutions.

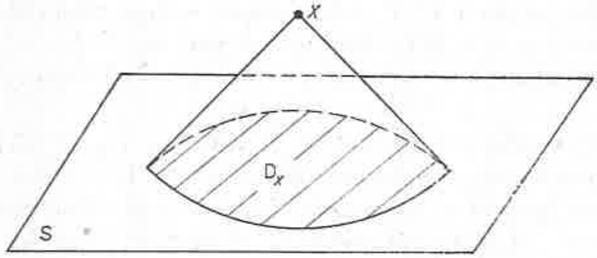


Fig. 2

In situations where rest-mass and/or scattering is present, Huygens' Principle will not hold; and in such cases the twistor translation appears to be not quite so neat. The presently-accepted way of dealing with rest-mass is to use functions of several twistors, and this approach brings with it a new way of looking at the classification of elementary particles [3, 4]. An example of scattering is provided by the "full" (i.e. non-self-dual) Yang-Mills equations; here also there has been progress in understanding the twistor translation [15—17].

4. Curved space-time

Up until now we have dealt only with twistors for flat space-time, but it is essential for the twistor programme that one should be able to translate curved space-time into the twistor language. In general, the presence of gravitation leads to scattering and the failure of Huygens' Principle, so in view of the remarks made in the previous section, we must expect matters to be less simple in curved space than in flat. Another (related) way of putting this is that there is in general no natural embedding of the space N of null geodesics into a three-dimensional complex manifold (cf. section 2).

This problem is one of the main areas of research in twistor theory. The most promising approach at present seems to be that of hypersurface twistors [2, 6]. The idea here is that if one selects a spacelike (or null) hypersurface S in space-time, then N can be naturally embedded in a twistor space T(S), which depends on S. The complex manifold T(S) contains, in its complex structure, information about the gravitational field "at S"; as one moves from one hypersurface to another, the structure of S "shifts". One interesting possibility that was recently raised is that hypersurface twistor theory might throw new light on the decomposition of fields into positive and negative frequencies (cf. section 2), and hence on the definition of particles in curved space-time [18].

There is a special class of curved space-time where, in a sense, Huygens' Principle *does* hold, and for which there is a very satisfactory and useful twistor description. These are the space-times with self-dual or anti-self-dual conformal curvature, i.e. where the Weyl tensor C_{abcd} satisfies

$$^1/_2\varepsilon_{abcd}C^{cd}{}_{ef} = \pm \mathrm{i} C_{abef}.$$

Because of the factor i in this equation, the only self-dual Lorentzian space is Minkowski space-time. Non-trivial self-dual spaces are either complex or have a non-Lorentzian signature (such as $++++$); both possibilities are of some interest [19, 20].

The remarkable feature of the twistor description is that Einstein's field equations "disappear" on translation into the twistor picture. Roughly speaking, one finds that self-dual solutions of the equations $R_{ab} = \lambda g_{ab}$ (for any desired value of the constant λ) correspond to holomorphic deformations of the "flat" twistor space T of section 2, preserving certain differential forms [21, 6, 22]. For some recent applications of this correspondence, see [23, 24]. It is worth emphasizing that a "direct" attack on the self-dual Einstein equations still leaves one with a nonlinear partial differential equation to solve [19]; the twistor approach eliminates the equation altogether.

In conclusion, a good deal more progress clearly needs to be made before the twistor programme can begin to fulfil its more ambitious aims. But enough remarkable features have already appeared to make a search for this progress seem well worth while.

References

[1] R. Penrose, in Magic without Magic, ed. J. R. Klauder, pp. 333—354 (Freeman, 1972).

[2] R. Penrose, in Quantum Gravity, eds. C. J. Isham, R. Penrose and D. W. Sciama, pp. 268 to 407 (Oxford, 1975).

[3] L. P. Hughston, Twistors and Particles (Springer Physics Notes no. 97, 1979).

[4] Advances in Twistor Theory, eds. L. P. Hughston and R. S. Ward (Pitman, 1979).

[5] S. A. Huggett and R. Penrose, Three Channels for the Box Diagram (preprint, Oxford, 1980).

[6] R. Penrose and R. S. Ward, in General Relativity and Gravitation, ed. A. Held, Vol. 2, pp. 283—328 (Plenum, 1980).

[7] M. G. Eastwood, R. Penrose and R. O. Wells, jr., Cohomology and Massless Fields (preprint, Oxford, 1979).

[8] R. F. Streater and A. S. Wightman, PCT, Spin and Statistics, and All That (Benjamin, 1964).

[9] R. S. Ward, Phys. Letters **A 61** (1977) 81—82.

[10] M. F. Atiyah, Geometry of Yang-Mills Fields (Scuola Normale Superiore, Pisa, 1979).

[11] M. F. Atiyah and R. S. Ward, Commun. math. Phys. **55** (1977) 117—124.
[12] M. F. Atiyah, N. J. Hitchin, V. G. Drinfeld and Yu. I. Manin, Phys. Letters **A 65** (1978) 185—187.
[13] J. Madore, J. L. Richard and R. Stora, in Feynman Path Integrals, pp. 304—344 (Springer Physics Notes no. 106, 1979).
[14] R. S. Ward, A Yang-Mills-Higgs Monopole of Charge 2 (preprint, Trinity College, Dublin, 1980).
[15] J. Isenberg, P. B. Yasskin and P. S. Green, Phys. Letters **B 78** (1978) 462—464.
[16] E. Witten, Phys. Letters **B 77** (1978) 394—398.
[17] G. M. Henkin and Yu. I. Manin, Twistor Description of Classical Yang-Mills-Dirac Fields (preprint, Steklov Institute, Moscow, 1980).
[18] R. Penrose, Some Remarks on Twistors and Curved-Space Quantization (preprint, Oxford, 1980).
[19] I. Robinson, in these Proceedings.
[20] G. Gibbons, in these Proceedings.
[21] R. Penrose, Gen. Rel. Grav. 7 (1976) 31—52.
[22] R. Ward, Self-Dual Space-Times with Cosmological Constant (to appear in Commun. math. Phys.).
[23] N. J. Hitchin, Math. Proc. Cambridge philos. Soc. **85** (1979) 465—476.
[24] M. F. Atiyah, Green's Functions for Self-Dual 4-manifolds (peprint, Oxford, 1980).

Address: Prof. Dr. R. S. Ward
Dept. of Mathematics
Trinity College, Dublin
Ireland

Black Holes, Singularities, and Topology

H.-J. Seifert (Hamburg)

A friendly warning: This is a written version of a lecture for a larger auditory and lacks in completeness and precision.

Two types of references have been preferred to "original papers" published in journals (the authors might forgive me):

- *Textbooks which are still up to date and review articles in the Einstein centenary volumes (comprehensive and well readable introductions into the subject with quite complete bibliographies),*
- *papers contributed to the discussion groups of the conference (to stimulate these discussions, with the risk that the reader who did not join the conference has difficulties to get them and the risk that some claimed results turn out to be wrong).*

One might wonder why "black holes" are scarcely explicitly mentioned in this paper. Now, exact solutions, astrophysical evidence, numerical calculations, and quantum effects had been covered by other lectures; so it was left to me to discuss the connection between black holes and singularities.

My aim is to sketch the present state of affairs and to suggest what problems should be attacked next. Doing so I shall not try to hide my personal biased views.

0. Introduction

The title (given by programme committee) could have been formulated more seriously, e.g.: "Recent progress in qualitative fundamental theory of gravitational fields" or "Global properties of space-time". There was good reason for not doing that. The recent progress is slow; since the last GR conference several nice papers have appeared but nothing really exciting happened. The non-existence of some results might be as important as the existing papers.

On the other hand, "black holes, singularities, and topology" are still thrilling objects. You can tell even non-physicists about them (the imperfection of the theory, the lack of details and precise theorems turn out to be virtues for such a conversation), and they will listen to you, because these three topics are deeply connected with a very old problem:

The physical cosmos generally was considered being *finite* in space and time (the farther you go back in occidental history, the tinier it was); the geometrical background was *infinite* Euclidean space. What happens beyond the border where physics stops and geometry goes on? and: what was before physics started? All the attempts failed to construct physical, philosophical or theological frame works in which these two questions are convincingly either answered or excluded.

General relativity is the first theory that brings together again the geometrical and the physical cosmos:

> A finite time can be related with *singularities*, a finite space can be related with a compact *topology*,

and which allows to construct models for the world in which the two questions from above turn out to be improperly posed.

Admittedly, they are still asked (compare the history of the singularity problem, beautifully described in [3]), global relativity is still full of belief, hope and "theology".

1. Singularities: the positive (physical) aspect. (Singularities as dramatic events in the life of our cosmos)

1.0. Why does one expect singularities?

Gravitation is a *contracting* interaction (nonlinear hyperbolic field equations), hence one expects crushes, caustics, poles of densities and similar events as one knows from hydrodynamics. Gravitation is a *universal* interaction (the field is the background geometry itself), hence one does not simply have localizable irregularities of the field on a still regular background. Instead, the whole description by the theory comes to an end; *incompleteness* rather than discontinuity is the expected manifestation of singularities.

1.1. Does a prediction of singularities have to be taken serious?

People agree that near physically relevant singularities of a general relativistic space-time the classical theory becomes invalid and has to be replaced by some quantum theory. They do not agree, whether (the still not existing) quantum gravity would remove the singularities. But even if it would, the situation near big bang or star collapse in the region, which has to be classically described, becomes so uncomfortable that — from the macroscopic view — some densities are practically infinite. Singularities of General Relativity could indicate extreme situations of a revised theory (it is easier to obtain criteria for infinities than for "very large" finite values).

1.2. Do singularities present a convincing picture of past origin and of final end?

Even after having removed god as an explicitly used ingredient of cosmology there is such a lot of belief in diverse principles left over that one could not convince all relativists with one picture. But besides the obvious requirement: agreement with observations, there is at least one important demand: in view of the fact that astrophysical objects grew old and have a finite lifetime as a result of increasing entropy and dissipative processes, past and future borders of the cosmos should be related with that process of aging, the picture of the collapse should explain why the history of objects comes to an end within finite time. This is expressed in Penrose's [5]

> *Time-Arrow Hypothesis* (TAH):
>
> Initially (big bang) the entropy of the geometrical/gravitational field should be minimal and becomes infinite in final collapses (and — as a measure for that entropy —the Weyl curvature $C^{\alpha}{}_{\beta\gamma\delta}$ goes from 0 to ∞).

2. Singularities: the negative (mathematical) aspect. (Singularities as breakdown of some space-time structure)

2.1. Classification (corresponding to the hierarchy of structures in General Relativity one can find a first subdivision of singularity types)

2.1.1. Non unique evolution (field equations)

The existence and uniqueness theorems (available at present time on the initial value problems for Einstein's field equations) require slightly[1]) higher smoothness assumptions than one needs for Lorentz spaces acceptable as space-times. Hence, the following types of singularities cannot be excluded:

(i) a solution of Einstein's field equations which cannot be extended as a solution but as a regular Lorentz manifold,

(ii) a solution which can be extended as a solution in different ways (more precise: identical initial data develop into not even locally isometric domains of dependence).

I do not know of any example of these types. Type (ii) is quite unlikely in view of the fact (Hawking & Ellis [1] proposition 7.5.2.) that two such branches *both* have to be coarse (hence analoga to the standard example: $\Box u + 3\left(\frac{\partial u}{\partial t}\right)^{2/3} = 0$; $u \equiv 0$ and $u \equiv \Theta(t_0)\,(t - t_0)^3$ with a branch at $t = t_0$ do not exist). For type (i) the gap between the smoothness requirements is smaller than for (ii) and more likely to be filled in near future.

---

[1]) It is difficult to give a simple but precise measure for this gap, as in the initial value theory Sobolev classes W^k are used rather than C^k.

2.1.2. Discontinuities (metric structure)

Given a smooth manifold, it is possible to consider fields (including the metric tensor field) which are discontinuous. On the other hand, in view of the fact that space-time is a manifold which — together with the metric — develops in time it is hard to see what the smoothness of a manifold with a discontinuous metric physically means. (A field theorist paints a picture on an already given — usually flat — canvas, the general relativist knits a sweater producing the pattern together with the background). For certain idealized situations (pointlike test particles, impulse shock waves, surface layers) one uses manifolds with some discontinuities. Unfortunately the usual methods of treating these formally singular situations cannot be easily translated to General Relativity:

— *Distributional* differential geometry is not fully developed and the accepted junction conditions do not go down to those low smoothness classes one wants to consider in some cases (cf. [4], sect. 5.3 (i); [2] p. 551f),

— *conserved quantities* are not easy to obtain on a space-time without symmetries. *Surface integrals* around the discontinuities suffer from the difference of covariant and partial derivatives; nevertheless, they had been used successfully for a long time (e.g. Einstein & Grommer [6] for the point particles).

2.1.3. Incompleteness (affine structure)

As explained in sec. 1.0. this is the type of singularities one expects. There is a complicated hierarchy of inequivalent incompletenesses ([8], sect. 5; [3], sect. 42). Fortunately, the Penrose-Hawking singularity theorems as well as the explicitly known examples have an incompleteness which is natural from the physical point of view ([1], p. 258): inextensible causal geodesics with finite (affine) length.

There are, however, examples of singularities (e.g. Reißner-Nordström) where no timelike (only null-) geodesics are terminating, in physical terms: all freely falling particles avoid falling into the singularity. It may be conjectured that a suitable version of Cosmic Censorship will forbid these "repellent" singularities.

2.1.4. Acausality (conformal structure)

The hierarchy of causal requirements is even more impressive than that for the completenesses (cf. [9, 10]). In context of singularity discussions one can come along with four causality classes:

(i) Non chronology: closed timelike worldlines exist.

(ii) Stable causality: a time function exists ([1]; prop. 6.4.9.), or: causal infinity can be incorporated into the causal ordering of space-time ([11]).

(iii) Global hyperbolicity: the whole spacetime is the development of initial data on a "Cauchy surface" S, or: the set of causal curves connecting any pair of events is compact ([1], sect. 6.6.).

(iv) Asymptotic simplicity (not a purely causal, i.e. conformal, requirement!) ([1], sect. 6.9.).

2.2. Description (singular boundaries)

A lot of different methods to attach the "endpoints" of incomplete non-extensible curves had been proposed in the past (a comprehensive survey is given in the diploma thesis of F. Müller [12]):

(i) The g-boundary (Geroch, Hawking) collects all endpoints of incomplete geodesics with the structure imposed by the initial data for the geodesic equation. Many examples, partly given by Geroch himself, show that for general Lorentz spaces the g-boundary is unsatisfactory. It may well be (and I believe it) that for the generic collapse singularity this boundary will have a renaissance (see sect. 4.4.).

(ii) The b-boundary (Schmidt, [1], sect. 8.3.) has a very elegant mathematical formulation, and despite the exceeding difficulties in investigating it even for the simplest special space-times it was considered until recently as the foremost description of singularities. But then it turned out that (Bosshardt, Johnson, [3], sect. 5.2.) this boundary is not reasonable unless it is modified (which lowers the elegance and enlarges the problem in calculating it).

(iii) The c-boundary (Geroch, Kronheimer & Penrose; Seifert; Budic & Sachs [1], sect. 6.8.; [3], sect. 4.2.) is comparatively simple to calculate (the c-boundaries of practically all explicitly given space-times are known) and describes at least that aspect of singularities which is disputed most (causal interaction, cosmic censor, horizons or mixmaster in the big bang?).

3. The character of singularities

3.1. Incompleteness is a general feature of space-time

The famous Penrose-Hawking singularity theorems (1965—1970) show that realistic space-times are generally (non extensibly) incomplete. To be a little more precise, the conditions

(i) a quite weak *energy* and generality condition (non-negativity of energy density which leads to gravitational attraction and *focussing*),

(ii) an *initial* geometric arrangement ("trapped set" which one expects to occur in collapse as well as in an expanding cosmos),

(iii) *causality* (which is a requirement if one applies certain theorems to show that extremal geodesic arcs which are *not* allowed to have *focussing* (conjugate) points exist)

are shown to contradict

(iv) *non-singularity* (completeness of all causal geodesics).

During the last decade only minor improvements of these theorems have been found.

Tipler gave evidence ([3], sect. 5.1. (ii)) that incompleteness is not prevented by acausality (i.e. (i), (ii) solely contradict (iv)). The omission of the merely technical assumption (iii) is plausible in view of the feeling that the focussing caused by (i) gives rise to the bounds on the length of geodesics; acausality at best repeats focussing rather than prevents it.

Beem and Ehrlich [13] generalized the initial situation (ii) to

(ii') Existence of causally disconnected sets: All causal curves between two sets A, B $(A \subset I^+(B),\ B \subset I^-(A))$ meet a compact set K, see fig. 1.

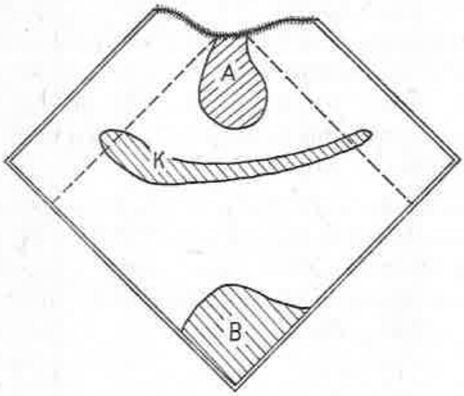


Fig. 1. Two non compact sets A and B causally separated by a compact set K.

3.2. The geometric nature of singularities

A singularity which has been developed from a regular situation should display in its geometry why the incomplete world lines do have finite lifetime. In other words, the curvature which measures the deviation from the flat geometry, strength of the gravitational field (Weyl tensor), and the matter density (Ricci tensor) should diverge sufficiently.

The following beautiful theorem presented by Isenberg and Clarke [14] at this conference is likely the most important result in singularity theory obtained in the last few years:

Curvature Singularity Theorem (CST)

Let M be the development of some initial situation (in other words: let M be globally hyperbolic), and let γ be a world line (causal curve) of finite (affine) length, then

— either the curvature $R^{\alpha}_{\beta\gamma\delta}$ is discontinuous (*curvature singularity*) or algebraically special (violation of generality) along γ,

— or M can be *extended* to a space-time $\tilde{M}$, which contains the endpoint of γ as a regular interior point, the boundary ∂M in $\tilde{M}$ being a Cauchy horizon generated by null geodesics.

It seems that the astrophysicists now can forget the jocose incompletenesses cut and sawn by the space-time tailors (for an example see fig. 2). Although, this theorem is not the ultimate answer, because there are reasons for believing that discontinuity of the curvature is not a guarantee that the geometry becomes sufficiently disastrous to bring all physics to a definite end, see sect. 4. One needs lower bounds for the rate of divergence of the curvature, and today we only know upper bounds (Tipler [3], sect. 5.4. (iii)).

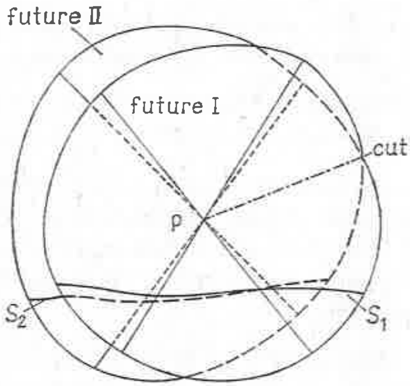


Fig. 2. A quasiregular singularity: double covering of 2-dim. flat space after one point p removed (In a "picture" one has a cross-over along the "cut", which does not affect the internal geometry). Branch-point p cannot be put back as it would possess two past- and two future-nullcones. If one poses an initial value problem on $S_1 \cup S_2$, two separated sheets of Minkowski space, hence a disconnected but regular space-time will develop.

4. Cosmic Censorship or whether singularities interact with the regular outside

4.1. Preliminary versions of the conjecture

The majority of relativists believes (or simply hopes?) in

> *Cosmic Censorship* (CCH): No (then called "naked") singularities can be seen by outside observers.

To find arguments for or against this hypothesis was the key issue in the singularity and black hole theory during the last decade.

The formulation of CCH as given above is misleading in at least two points:

(i) Naked Singularities are not rejected because one has a horror of glancing at singularities; the initial big bang singularity is in the causal past of all of us "observers", but nobody considers this as a counter-example to CCH. What one really likes to exclude is an interaction with regular regions: the region, where (classical) General Relativity holds, should be, in concordance with the ideology of classical deterministic physics, the time evolution of some regular initial data surface S. This (principal) predictability breaks down if singularities act into the future of S without leaving a trace of this influence on S itself or without being incorporated into the theory (a distribution calculus or some conserved quantities or integral formulas might allow to remove some uncertainty). Briefly stated:

> *Strong Cosmic Censorship* (SCCH): That part of space-time which is describable by Einstein's theory is also predictable by Einstein's theory.
>
> In other words: the whole (regular) space-time M is the evolution of the data on some Cauchy-surface S; M is globally hyperbolic.

The cosmic censor, who was originally only involved in the collapse singularity and far distant observers, now regulates the interior of black holes (no interaction between singularity and observers diving into the black hole) and, in passing, forbids timelike parts of regular infinity (which, in fact, could only occur for unreasonable matter densities or nonvanishing cosmological constant).

(ii) There are a lot of special solutions *violating* CCH or SCCH. Until now the discussion has been focussed on ruling out these examples as non-relevant. The time has come to collect all these scattered results and arguments to formulate some more precise versions of CCH which are not disproved by known counter-examples and at least have a chance to be proven or disproven within the next decade.

4.2. The challenge by counter-examples

Most discussed in this context is the "shell crossing" singularity ([15], [8], [3], sect. 6.2. (ii); [7], several articles) of spherically symmetric collapsing perfect fluid ball.

(i) Shell crossing has analogues in Newtonian hydrodynamics. But: big bang has a Newtonian analogue, too and is still a most serious singularity.
On the other hand: big bang is an *initial* singularity, collapse a *final* one where one expects according to TAH (sect. 1.2.) that gravity (Weyl tensor) dominates matter (Ricci tensor).
Admittedly, shell crossing singularities are not the inevitable and definite end of the collapse which, in fact, continues and later on forms a horizon; nevertheless, the singularity influences the regular outside region and, in contrast to Newtonian hydrodynamics, produces a singularity of the gravitational field.

The fact that in hydrodynamics shell crossing can be removed leads to the next suggestion:

(ii) Shell crossing only occurs as one uses an unrealistic equation of state; perfect fluid cannot be a good approximation if the shear of the matter flow tends to infinity, and therefore, no matter how small the viscosity coefficients are, the viscosity stresses blow up, or more likely, damp the shear and prevent the shell crossing.

> *Hydrodynamic singularity conjecture* ([7] comments by Hawking and Seifert). If one uses a description of state (like viscosity) which in Newtonian hydrodynamics does not allow infinite densities, naked singularities[1]) will be prevented in almost spherical collapse.

Viscosity damps out the relative motion of flow lines which is started by the initial arrangement (in the shell crossing case: outer shells shrink quicker than some inner ones); dynamical viscosity will prevent infinities of shear, bulk viscosity might retard the growing of convergence but cannot stop it as gravitation dominates finally and leads to a hidden collapse:

> *Hoop conjecture* (Thorne [19]). If a suitable amount of matter is concentrated in a region with a given diameter (in all directions, a hoop surrounds the region), a horizon (black hole) forms.

Before I continue (in the next two sections) the discussion opened by the shell crossing, I would like to mention an example which has attracted much less attention but is even more puzzling: Winicour, Janis, and Newman [20] found a family of static solutions (Schwarzschild plus an arbitrary small quadrupole moment) with a singular pointlike event horizon; as a matter of fact, it does not violate SCCH (the singularity is nowhere timelike), it marginally violates CCH (disproving the wording "Strong-" and "Weak-"CCH) and heavily violates concepts of stability of the Schwarzschild black hole which arise from a naive interpretation of Penrose's first singularity theorem. Recently, Kates [21] has revisited these solutions and argues that singular event horizons can occur also in the dynamical process of collapse.

4.3. Retreat to the pure vacuum

The other, undoubtedly simpler direction to escape the problems of shell crossing matter is to restrict the investigations to pure vacuum solutions in order to find those singularities which are really gravitational and not generated by an accidental arrangement of matter ([4], conjecture 5.1.; [3], sect. 6.2.(ii)):

> *Vacuum Cosmic Censorship* (VCCH): Asymptotically flat vacuum initial data cannot develop into naked singularities.

[1]) With the methods of Lifshitz and Khalatnikov, Grishchuk has constructed [16] a stable solution for viscous fluid with a non-spacelike hence naked singularity. One more reason to clarify, whether the "Russian stable singular solutions" can be strictly established or definitively disproved (cf. the dispute between Barrow & Tipler [15] and Belinsky, Khalatnikov, Lifshitz [16]).

Three papers presented at this conference signalize a progress into this direction:

Christodoulou and O'Murchadha [22] show that the evolution of asymptotically flat vacuum data not only covers some neighbourhood of the initial surface but contains at least one hyperboloid (the closer the data are to the flat metric, the larger is the range of existence of the solution).

Moncrief and Eardley [23] give arguments that VCCH would be a consequence of a global existence theorem in the "slice gauge" ($t = t_0$ surfaces are extremal spacelike hypersurfaces[1]); global means: for $t_0 \to \infty$) for which they have already obtained an analogue for certain compact slices.

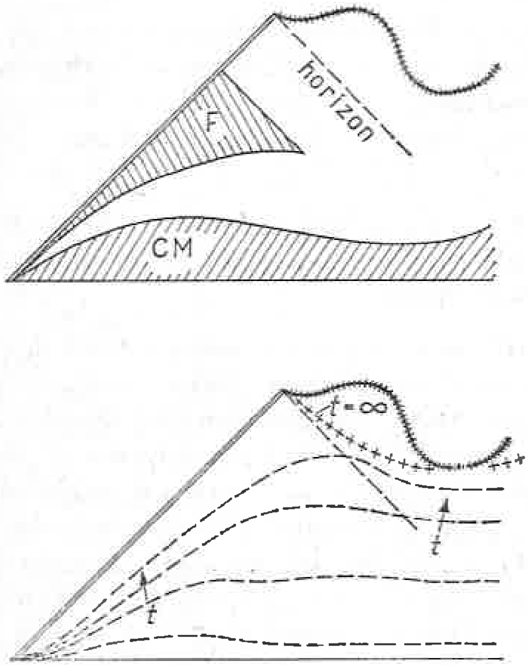


Fig. 3. The regions where the existence of a solution for initial value problems has been shown (F: Friedrich, CM: Christodoulou and O'Murchadha) and where one hopes to show it (VCCH; the dotted line marks the slices t = const in Eardley, Moncrief).

Friedrich [24] proved that, for general data posed on a regular $\mathscr{I}$, vacuum solutions exist. This is the first theorem proving rigorously the existence of dynamical vacuum solutions which are (weakly) asymptotically simple and, for that matter, that the set of space-times for which the (original) Cosmic Censorship is a sensible requirement is not empty. Furthermore, the domain of existence fits nicely together with the ones of the theorem and the conjecture mentioned above, see fig. 3.

---

[1]) The existence of which still has to be shown, despite the universal belief in them.

No doubt that the proof of VCCH would be a great success. But I do not share the view of Horowicz ([4], [7]) that this would almost settle CCH. Let us consider the following example:

> An axisymmetric star with an exterior vacuum which is (at least for $t \to \infty$) the Tomimatsu-Sato (TS) solution might collapse starting at a regular initial situation. Such a star will uncover the naked singularities of the TS-solution outside the horizon, hence undoubtedly developing a naked vacuum singularity violating CCH. On the other hand, the TS vacuum (without the star) would possess the singularity from the beginning; we cannot find regular initial data and therefore have no violation of VCCH. I do not know any arguments within local physics which say that TS-solutions are fundamentally inferior to Kerr solutions as stationary endstate solutions: the characterizing theorem for Kerr-fields (Israel-Carter-Hawking-Robinson) already assumes what one would like to prove here: no singularities between black hole and $\mathscr{I}$.

If one investigates properties of the collapse (CCH was especially formulated for this situation) by throwing away the matter ball, the "reason" for the outside vacuum field, one may be left with an unreasonable field.

4.4. Might curvature singularities not be singular enough?

The last class of arguments in the "shell crossing discussion" that I want to mention is concerned with *strength.*

If matter density blows up on a sphere, the surface density remains finite; in other words, the *rate of growth is smaller* than in the pointlike collapse. The more *mathematical* approach is to go back to the treatment of discontinuity hypersurfaces (sect. 2.1.2.). Whether *physical* processes necessarily come to an end is a question of how strongly the gravitational *tidal forces* will damage a body encountering the singularity.

One has to specify (A) the structure of the body (physical system) and (B) what one considers as the "action" of the force.

A (i) Dust cloud (world lines are geodesics),

(ii) rigid body (say: world lines have constant Fermicoordinates based on some "central line"),

(iii) elastic body (something between (i) and (ii)).

In all cases the system should better be treated as finitely *extended* (not, as is usually done[1]), as an infinitesimal test body), as one expects the curvature to be very drastically over any characteristic length of the body. This implies that one has to calculate some mean values by integration.

[1]) Ellis and Schmidt ([7], p. 994) apply the calculus for gravitational wave detectors ([2], sect. 37) on the Schwarzschild singularity: Mashhoon [25] calculates the tidal force components along world lines with Taylor's formula.

B (i) The system consists of pointlike atoms (tidal forces are obtained by integration of the tidal force producing components of the Riemann tensor along connecting arcs between the atoms).

(ii) The atoms are smeared out (integration over the bulk volume).

(iii) Energy-momentum transfer, work (integration of force along the central line).

To give the notion:

> *Tidal stress singularity:* The tidal forces become infinite if one approaches the singularity,

and the conjecture:

> *Tidal stress Cosmic Censorship:* All tidal stress singularities are hidden behind horizons,

a precise meaning one has to specify the structure of the detector. Some difficulties:

- For sufficiently accelerated bodies the total energy transfer can be kept finite, B (iii) seems sensible only for freely falling bodies (geodesic central lines); cf. Ellis' and Schmidt's result that an accelerated detector could escape destruction before reaching the singularity ([7], p. 996).
- Which curves one should take as connecting lines in B (i) (for spacelike geodesics, all tidal forces in the sense of B (i) in the shell crossing solutions remain finite, for connecting lines with $r = \text{const}$ they can diverge):

For *geodesic* world- and connecting-lines, the shell crossing singularities in all the models discussed above have finite tidal forces, the Schwarzschild-Kruskal singularity has infinite forces.

Yodzis suggested ([8], sect. 8) a related type of a "strong singularity" which is geometrically formulated and circumvents the ambiguities discussed above. The limits of the Christoffel symbols in normal coordinates have to be evaluated. (The Γ's are 1^{st} order quantities similar to the integrals over the tidal force producing components; this "*geodesic singularity*" is comparable with tidal force singularity in the sense of A (i), B (iii)).

After arriving at a satisfactory definition of "physically strong singularities" or "breakdown of C^1-structure singularities" one still has to find a relation to the singularity existence theorems (analogue to Clarke's results for curvature singularities).

5. Topology

5.1. The global manifold

Once when relativists discovered topology, they tried to find out *how many* different topological (global) structures space-times can have (infinitely many). Today, one tries (quite successfully) to reduce all these possibilities down to four: $C^3 \times \mathbb{R}$ ($C^3 \cong \mathbb{R}^3$ or S^3 or $\mathbb{R} \times S^2$ or $\mathbb{R}^2 \times S^1$; S^k being the k-dimensional sphere).

An early theorem stated:

> A globally hyperbolic space-time (admitting a Cauchy surface C) is homeomorphic to $C \times \mathbb{R}$.

For cosmology this can be used to show:

> A nearly spatially homogeneous, simply connected space-time is homeomorphic to $\mathbb{R}^4$ or $S^3 \times \mathbb{R}$.

For the study of bounded astrophysical systems one obtains the following results (Tipler; Brill & Lindblom [26]).

(i) Any non-singular, weakly asymptotically flat and empty, causally regular and generic space-time is globally hyperbolic,

(ii) if, in addition, the past part is asymptotically simple, space-time is homeomorphic to $\mathbb{R}^4$.

5.2. The local manifold

Most (classical) physicists take the manifold (or even the Euclidean) structure for granted. After Kronheimer and Penrose promoted causal structure as a (or *the*) basic structure and obtained the topology from the causal ordering ([1], p. 196), it was natural to ask, whether the geometrical properties of space-time could be described in terms of causality. The basic properties (connected, countable, etc.) can be quite easily reformulated; but a convincing characterization of a manifold in terms of general topology has not been found till now.

One might conjecture that manifolds are those topological spaces which are particularly *homogeneous* [27], but only for very low dimensional manifolds one has found theorems:

(i) Any (linearly) ordered, connected, separable set is homeomorphic (in the ordering topology) to an interval of $\mathbb{R}$; this gives a nice description for curves as "world lines".

(ii) Any connected, locally connected, non-degenerate metrizable space which is

- compact and separated by any pair of points is $\cong S^1$,
- separated by any homeomorphic map of S^1 but not by a pair of points is $\cong S^2$ (Kline).

This gives a comparatively complicated characterization of the set of all directions (where light rays come from; the separation property says that silhouettes are closed curves).

Using these theorems one can construct local parts of space-time in "pine-tree" coordinates: null cones (direction S^2) $\times$ (null ray $\mathbb{R}$) along a world line $\mathbb{R}$. $(S^2 \times \mathbb{R} \cup \{p\}) \times \mathbb{R}^3 \cong \mathbb{R}^4$.

For a characterization of differentiability in a similar way I cannot present even a vague conjecture.

6. What to do till the next GR conference

I expect that some progress can be obtained in the following topics:

Differential geometry:

— Study of geodesics near singularities (1st order singularities), sect. 4.4.;
— Existence of extremal spacelike hypersurfaces, sect. 4.3.;
— Hoop conjecture, sect. 4.2.;
— Treatment of discontinuities, sect. 2.1.2.

Field equations and initial value problems:

— Vacuum Cosmic Censorship, sect. 4.3.;
— Belinski-Khalatnikow-Lifshitz singular solutions;
— Viscosity and anisotropic collapse, sect. 4.2.

In any case, a closer familiarity with the theory of nonlinear field equations is a demand on any "global relativist". Admittedly, partial differential equations are more formalistic and less enjoying than qualitative differential geometry. Brilliant results in global geometry have been presented in 2-page papers (Penrose's singularity theorem, Hawking's monotony of black hole area), but for results on differential equations you need at least 20 pages. But lack of appeal is not lack of importance. To my opinion, one often underestimates the role of the full field equations, e.g. by the statement that singularities in a solution will be avoided unless they are forced by a few weak consequences of the full field equations ([7] Hawking's comment). It seems to me of minor importance to look for further definitions unless one has a chance to use them in theorems (more boundaries of space-time; more definitions of "black holes" in general spaces, etc.).

References

[1] Hawking, S. W.; Ellis, G. F. R.: "*The Large Scale Structure of Space-Time*". Cambridge University Press (1973).
[2] Misner, C. W.; Thorne, K. S.; Wheeler, J. A.: "*Gravitation*". San Francisco (1973).
[3] Tipler, F. J.; Clarke, C. J. S.; Ellis, G. F. R.: "*Singularities and Horizons*" in: General Relativity and Gravitation, ed. A. Held, New York & London (1980).
[4] Geroch, R. and Horowicz, G. T.: "Global Structure of Spacetimes", and
[5] Penrose, R.: "Singularities and Time-Asymmetry", both articles in "*General Relativity, an Einstein Centenary Survey*", ed. S. W. Hawking and W. Israel. Cambridge (1979).
[6] Einstein, A., Grommer, J.: „Allgemeine Relativitätstheorie und Bewegungsgesetz". Sitz.-Ber. Preuss. Akad. 1927, I (reprinted: Berlin 1978).
[7] GR 8 Symposium on Singularities, organized by G.F.R. Ellis, *Gen. Rel. Grav.* **10**, 959 to 1069 (1979).
[8] Seifert, H. J.: "The Causal Structure of Singularities" in: *Differential Geometrical Methods in Mathematical Physics*. Springer Lecture Notes in Mathematics 570, Berlin (1977).

[9] Carter, B.: "Causal Structure in Space-Time". *Gen. Rel. Grav.* **1**, 349—391 (1971).

[10] Hawking, S. W.; Sachs, R. K.: "Causally Continuous Space-Times". *Commun. math. Phys.* **35**, 287—296 (1974).

[11] Seifert, H. J.: "The Causal Boundary of Space-Time". *Gen. Rel. Grav.* **1**, 247—59 (1971).

[12] Müller, F. H.: „*Raum-Zeit-Ränder*". Diploma thesis, Göttingen (1979).

[13] Beem, J. K.; Ehrlich, P. E.: "Singularities, incompleteness and the Lorentz distance function". *Math. Proc. Cambridge Phil. Soc.* **85**, 161—178 (1979).

[14] Isenberg, J.; Clarke, C. J. S.: "On Space-Time Developments of Cauchy Data". Contributed paper for GR 9 (1980).

[15] Yodzis, P.; Seifert, H. J.; Müller zum Hagen, H.: "On the Occurrence of Naked Singularities in General Relativity I/II". *Commun. math. Phys.* **34**, 135—148 (1973); **37**, 29—40 (1974).

[16] Grishchuk, L. P.: "Physical Singularity in a Space filled with a non Ideal Medium". JETP Letters **6**, 184 (1967).

[17] Barrow, J. O.; Tipler, F. J.: "Analysis of the Generic Singularity Studies by Belinskij, Khalatnikov, and Lifshitz". *Phys. Reports* **56**, 371—402 (1979).

[18] Belinskii, V. A.; Khalatnikov, I. M.; Lifshitz, E. M.: "On the Problem of the Singularities in the General Cosmological Solution of the Einstein Equations". *Phys. Letters* **77A**, 214—216 (1980).

[19] Thorne, K. S.: "Nonspherical Gravitational Collapse", in: *Magic Without Magic: John Archibald Wheeler*, ed. Klauder, San Francisco (1972).

[20] Winicour, J.; Janis, A. I.; Newman, E. T.: "Static, Axially Symmetric Point Horizons". *Phys. Rev.* **176**, 1507—13 (1968).

[21] Kates, R. E.: "*Possibly Singular Small-Amplitude and Low Frequency Limits for Radiatable Fields in Gravitational Collapse*" (preprint Harvard-Smithonian 1980).

[22] Christodoulou, D.; O'Murchadha, N.: "*The Boost Problem in General Relativity*". Contributed paper for GR 9 (1980) submitted to Comm. Math. Phys. (München MPI/Astro 1980).

[23] Eardley, D.; Moncrief, V.: "*The Global Existence Problem and Cosmic Censorship*". Contributed paper for GR 9 (preprint Yale 1980).

[24] Friedrich, H.: "*On the Regular and the Asymptotic Characteristic Initial Value Problem for Einstein's Vacuum Field Equations*". Contributed paper for GR 9 (preprint Hamburg 1980).

[25] Mashhoon, B.: "*On the Nature of a Spacetime Singularity*" (preprint Pasadena Inst. of Technol. 1980).

[26] Lindblom, L.; Brill, D. R.: "*Comments on the Topology of Nonsingular Stellar Models*", (to appear in the Taub Festschrift).

[27] Cannon, J. W.: "The Recognition Problem: What is a Topological Manifold". *Bull. Amer. Math. Soc.* **84**, 832—866 (1978).

Acknowledgement:

I wish to express my admiration for the splendid organization of the GR 9 conference and I am grateful for the kind hospitality I experienced in Jena.

Address: Prof. Dr. H.-J. Seifert
Hochschule der Bundeswehr Hamburg
— Fachbereich Maschinenbau —
Holstenhofweg 85
D - 2000 Hamburg 70
Federal Republic of Germany

Terrestrial and Planetary Relativity Experiments

R. D. Reasenberg and I. I. Shapiro (Cambridge/MA)

Abstract

The advance of technology has made possible the current activity in experimental relativity. Radio and optical ranging techniques provide interplanetary and lunar distance measurements with fractional uncertainties as small as 10^{-11}. The analysis of these data yields results consistent with general relativity; the fractional uncertainty in the confirmation of specific effects is as low as 10^{-3}. Future experiments may extend the range of effects checked and permit the accuracy of some tests to be increased by several orders of magnitude required to detect post-post-Newtonian phenomena.

1. Introduction

Physics, as a productive human enterprise, has been able to outshine many other activities because of the balanced interplay of theory and experiment. For a long time, general relativity was in danger of being an exception to this pattern, being solely the province of theorists. Today, at least, the experimentalist is able to make a contribution to this subject thanks to the availability of high technology.

We first discuss relevant examples of this technology and then review the status of terrestrial and planetary experiments to test general relativity.

2. Instrumentation

2.1. Ranging to Passive Targets

The first high technology applied to experimental relativity was radar. As early as 1946, radar observations of the Moon were possible using techniques that were an outgrowth of World War II developments. By 1964, echo delays from radar observations of planets were being made with accuracies sufficient to be of use even today in studying the dynamics of the solar system. With the Arecibo radar, these delays are now being made routinely with uncertainties of a few tenths of a microsecond, equivalent to range errors of a few tens of meters.

The most severe difficulty that one has in using these radar observations is not with their accuracy but with the uncooperative nature of the planets. They are not polished spheres. Typically, planets have topographic variations which modify the round-trip propagation time of the radar signals by up to a few tens of microseconds, as compared to the several-tenth microsecond uncertainties of the current radar measurements of echo delays. In addition to continental-size constructs, there are also mountains, craters, and rilles of various sizes which add topographic noise from this level down to the measurements uncertainties and below.

What can one do about such degradation of precise measurements? Our first approach was to use an *ad hoc* model for the surface heights on each planet. The general "rule of thumb" was: the more free parameters, the better the model. We currently use such a model with 123 free parameters for the surface of each inner planet. This model "absorbs" much of the topographic structure, yet a great deal remains. We also analyzed the topography statistically, modelling the surface heights as a correlated noise process. However, the resulting $\sim 20\%$ improvement in the accuracy of parameters estimated from the radar observations of Mars, for example, is not sufficient to justify continued development of this approach because of the availability of substantially better data, not yet fully exploited. A third approach is the utilization of auxiliary data, like the topographic map of Venus (Pettengill et al., 1980) obtained from the radar altimeter aboard the Pioneer 12 spacecraft in orbit about Venus. We are now applying the information in this map in our analysis of solar-system dynamics.

Lunar laser ranging is another marvelous example of modern technology. The several sets of retroreflectors on the surface of the Moon are by-products of the Apollo and Lunakhod programs. Each laser ranging system now in use consists of a laser on the Earth that sends light beams through a telescope to a retroreflector; the return beams pass through either the same or a different telescope to a detector. With the aid of electronics, the round-trip propagation time is measured very accurately, often to within a few tenths of a nanosecond. The McDonald Observatory has been involved in lunar laser ranging for over 10 years, longer than any of the others. Its laser transmits a $3J$ pulse, 20 times per minute. The receiver system on average detects one photoelectron per 10 pulses transmitted, even though the retroreflectors return 100 times as much power as does the illuminated surface of the Moon. A typical 15-minute session of observations yields a "normal point," an equivalent single measurement of the distance from the observatory fiducial point to a retroreflector, with an uncertainty of 10 cm. The uncertainty in instrumental calibration is about 5 cm.

2.2. Ranging to Active Targets

One way to circumvent the problems of planetary topography is to observe spacecraft instead. Spacecraft provide superb targets for studies of solar-system dynamics. However, their use, too, is not devoid of problems. One must make precise measurements and convert them to equivalent measurements made in vacuum; in addition, one must have an accurate model for the location of the end points of the propagation path. Spacecraft carry transponders which usually allow measurements of both a Doppler shift and an echo time delay. With current transponders, for an integration time of 1 minute, Doppler measurements have an uncertainty of about 3 mHz, or about one part in 10^{12} of the 2.3 GHz carrier frequency; however, since the radial velocity between Earth and spacecraft is usually of the order of $10^{-4}c$, the fractional accuracy of the Doppler-shift measurements is about one part in 10^{8}. The time-delay measurements now have uncertainties as small as 10 ns which is about one part in 10^{11} of a typical round-trip echo delay.

In discussing the location of the end points of the propagation path, it is useful to consider three separate classes of spacercaft. The first is the deep-space probe; a craft sent from Earth into solar orbit. A Mariner class spacecraft, with an area-to-mass ratio of about 0.1 cm^2/g, is a rather flimsy object compared to the natural solar-system bodies observed with radar. Solar radiation pressure can cause an acceleration of such a spacecraft of order 10^{-5} cm/s^2; even the solar wind produces an acceleration of order 10^{-8} cm/s^2. The Mariner class spacecraft carry active attitude control systems which tend to produce slightly unbalanced thrusts from their gas jets. These cause accelerations of the order of 10^{-6} or 10^{-7} cm/s^2, which tend to be stochastic and are therefore difficult to model accurately. To place these numbers in context, note that after 10 days of acceleration at 10^{-7} cm/s^2, the spacecraft is displaced far enough to produce a change in the time-delay observable of 2.5 μs compared to the 0.1 to 0.01 μs uncertainty of measurement. Clearly, this situation is undesirable.

An improvement is realized when the spacecraft is placed in an orbit around a planet; the effect of "reasonable" unmodelled accelerations can then be estimated on a time-scale comparable to the orbital period. Thus a planetary orbiter, whose orbital period is typically about one day, can be located much more accurately than can a deep-space probe, whose orbital period is comparable to a year. There are, of course, penalties to be paid for being close to a planet. Reflected and re-radiated solar energy adds a new dimension to the complexity of modelling radiation pressure. The atmosphere of the planet, for a spacecraft that dips near to the surface, may perturb the orbit importantly. The Pioneer 12 orbiter of Venus, discussed above, experienced changes of orbital period of as much as 30 s during a single passage through periapsis. Perhaps most importantly, the gravity field of the host planet is not spherically symmetric and the higher-order spherical harmonics collectively perturb the trajectory of the spacecraft. Nonetheless, a planetary orbiter is preferable to a deep-space probe, if the object is to locate the spacecraft. From this point of view, the best spacecraft currently possible — the third class — are planetary landers. Gone are all the standard problems with spacecraft in flight. Arising are only a few new, much smaller problems.

2.3. Clocks

We now discuss devices used to measure time. For relativity experiments, hydrogen masers are now preferred because of their greater stability over time scales of hundreds to tens of thousands of seconds. Substantial further improvements in these clocks seem feasible, especially with operation at low temperatures if suitable wall materials can be developed.

Another clock, under development by Turneaure (Stein and Turneaure, 1975) is a superconducting cavity stabilized oscillator. With this technique, a superconducting cavity is illuminated with a signal from a voltage-controlled oscillator. The deviation between the oscillator frequency and the cavity resonance frequency is forced to zero by the action of a closed loop feedback system which drives the oscillator to the cavity frequency. With this instrument, an Allan variance close to 10^{-16} is being obtained for intervals of about 100 seconds (Turneaure, 1980).

A different type of cavity stabilized oscillator, based on a cavity without metallic walls, is also being developed (Braginsky, 1980). Made of a piece of sapphire in a ring configuration, the principal advantage of this cavity is an extremely low coefficient of thermal expansion. If temperature variations are the principal source of drift, then use of sapphire at liquid helium temperatures will result in a fractional change in frequency of about $5 \times 10^{-12} K^{-1}$. The Q of such a cavity has not yet been measured, but based on laboratory measurements of the power factor of this material at 3 cm wavelength, the Q is expected to be of order 5×10^8.

3. Experiments

The solar system is probably the best understood of the laboratories in which experiments in relativity are feasible to perform. The next best is now provided by the binary pulsar, PSR1913+16. With any such laboratory, there are always aspects of the system not incorporated in the model used in the data analysis. In fact, with the solar system, there are many aspects of the dynamics that we know about and do not incorporate in our model because unknown physical parameters associated with these aspects would also have to be estimated. If we tried to incorporate all of these, the number of parameters to be estimated would become "astronomical" and the estimator, degenerate. Model error is guaranteed to produce biased parameter estimates. So, one is caught between degeneracy and model error. Our procedure has been to start with all of the parameters that appear to be of importance, and then to do a large number of numerical experiments incorporating first one, then another, and then some combination of the smaller effects in an attempt to bound the errors that result from the model errors.

3.1. Principle of Equivalence

The principle of equivalence is often considered the foundation stone of general relativity. As early as the 5th century, this principle was a topic of discussion by Ioannes Grammaticus (Cohen and Drabkin, 1948) who asserted its correctness. His approach, of course, was somewhat different from that taken today. A number of laboratory experiments conducted in more recent centuries have yielded bounds on any possible violation of this principle. If we define ε as the deviation from unity of the ratio of gravitational to inertial mass, then the results of the experiment by Eőtvős were claimed to satisfy $|\varepsilon| < 5 \times 10^{-9}$; that of Roll, Krotkov, and Dicke (1964), $|\varepsilon| < 3 \times 10^{-11}$; and that of Braginsky and Panov (1971), $|\varepsilon| < 3 \times 10^{-12}$. However, for this last experiment, very few details have been published. The modern results imply that equal numbers of electrons and protons; neutrons; electromagnetic, strong, and weak binding energies; and kinetic energy all obey the principle of equivalence.

What do these experiments tell us about gravitational self-energy? Let us define the fractional gravitational self-energy, Δ, as

$$\Delta = M_s/M_r,$$

where $M_s c^2$ is the gravitational binding energy, or self energy, and M_r is the rest mass. If we consider a typical laboratory body, say a sphere of gold 1 m in diameter, we find that Δ is of the order of 10^{-23}; thus even the most precise principle-of-equivalence experiments performed in the laboratory could not begin to address the question of whether gravitational self energy satisfies this principle. One must consider much larger bodies since the gravitational self energy varies as the square of the scale size of the body.

Nordtvedt (1968) has considered the possibility of violation of the principle for astronomical bodies. In his analysis, Nordtvedt discussed the relation between the parameterized-post-Newtonian (PPN) formalism (Will, 1973) and η, a coefficient he introduced to represent any violation of the principle for gravitational self energy:

$$M_g = M_i(1 + \eta\Delta).$$

In this equation, M_g and M_i are, respectively, the gravitational and inertial masses of a body; $\eta = 0$ implies no violation.

Laser measurements of the Earth-Moon separation are sensitive to the difference between Δ (Earth) and Δ (Moon); fortunately these are very different. The observable effect, for $\eta > 0$, would be a displacement towards the Sun of the orbit of the Moon around the Earth. The corresponding change in the Earth-Moon distance would have a monthly period and an amplitude of about 8η meters. The ordinary, Newtonian gravitational effect of the Sun, however, also causes a monthly variation in the Earth-Moon distance whose phase is the same as for $\eta > 0$, but whose amplitude is 110 km! So one must, in effect, detect a small signal in the presence of a far larger one. Fortunately, the solar term can be determined to more than sufficient accuracy from other manifestations of the solar perturbations. Thus, from prior analyses of classical observations, its amplitude is known to within about 1 cm.

One of course employs a rather detailed and precise model to analyze the lunar laser ranging data. This model includes the effects of many perturbations on the orbital and rotational motion of the Moon such as the low-order spherical harmonics of the gravitational fields of the Moon and Earth and the tidal torque produced by the Earth. The model also includes the effects on the position of the observatory in space due to the irregularities of the position of the pole and the rate of rotation of the Earth.

From analysis of some 2000 of the lunar laser "normal points", obtained from about 5 years of observation, the principle of equivalence has been upheld. The estimate of η was consistent with zero. The two groups that analyzed the data assigned different uncertainties to their estimates: 1.5% (Shapiro et al., 1976) and 3% (Williams et al., 1976), the former being fourfold larger than the statistical error accompanying the estimate. We should emphasize that the allowance for systematic errors is somewhat subjective and depends on approximations made in the theoretical model, on trends discernible in the residuals, and on experience in the analysis of similar data.

Our MIT group has since incorporated the more accurate data gathered in the past five years and has also reduced the residuals substantially, principally by improving the model used for the rotation of the Earth. A new estimate of η should be published within a year.

Another system in which one could seek evidence for any violation of the principle of equivalence is the Earth-Mars-Sun-Jupiter system. We are analyzing a combination of radar and spacecraft data to estimate η for this system.

3.2. Gravitational Redshift

In their now famous redshift experiments, Pound, Rebka, and Snider used the Mőssbauer effect to obtain a γ-ray spectrum from an Fe^{57} sample with an extremely narrow line width. A similar crystal was used as a radiation absorber. The source and absorber were separated by a 26 m height and the predicted gravitational redshift was confirmed to within 1% (Pound and Snider, 1965).

More recently, a hydrogen maser clock was flown in a rocket to an altitude of 10,000 km to test the "redshift" prediction. The first-order Doppler shift of up to 2×10^{-5} had to be determined or cancelled to sufficient accuracy to observe the predicted gravitational effect of up to 4×10^{-10}. A clever arrangement of the tracking instrumentation provided cancellation of the first-order Doppler shift, except for that portion which resulted from the acceleration of the tracking station during the round-trip propagation of the signals to the spacecraft. The latter was easily included in the post-flight analysis. Similarly, the effect of the neutral atmosphere on the signal propagation was automatically cancelled to the extent that the atmospheric contribution was constant during the $\lesssim 0.1$ s round-trip propagation time from ground to rocket. A judicious choice of tracking frequencies and technique also provided virtual cancellation of the contribution of the ionosphere. The predicted gravitational redshift was confirmed to well within the uncertainty of 140 parts per million (Vessot et al., 1980).

What kind of experiment could be done to test the predicted gravitational redshift at a level of accuracy substantially higher than has been possible with a clock in a rocket? Clearly, such an experiment could be performed with a clock carried by a suitable spacecraft in solar orbit. With a perihelion distance of four solar radii, where the dimensionless gravitational potential is about 0.5×10^{-6}, such a spacecraft would carry the clock through a change in potential about one thousandfold greater than in the rocket flight. Unfortunately, there are serious, although not insurmountable, engineering and economic problems associated with such a spaceflight; a proposed mission, with the needed trajectory, is discussed in Section III 5.

3.3. Light Bending

The bending of light rays by the Sun was one of the early predictions deduced from the general theory of relativity. The predicted effect, 1''.75 for Sun-grazing rays, is multiplied by $(1 + \gamma)/2$ in the PPN formalism. Visual sightings of stars during a total solar eclipse were the first data to be used to check this effect. The most recent such observations were carried out by the Texas Mauritanian Eclipse Expedition of 1973. Although adverse weather conditions degraded the seeing and the accuracy of the results, the Texas team (1976) obtained $(1 + \gamma)/2 = 0.95 \pm 0.11$.

Substantially better accuracy can be obtained by the use of radio interferometry (Shapiro, 1967). Two or more separated antennas are used to receive signals from distant radio sources. The goal is to determine the apparent position of one source relative to another as a function of their angular separation from the Sun. In very-long-baseline interferometry, a hydrogen maser or other very stable clock is used at each antenna to convert the incoming microwave signal to a video signal which is subsequently recorded along with timing data on magnetic tape. The magnetic tapes from the antenna sites are brought together and cross-correlated to construct the fringes for the interferometer. In connected-element interferometry, a common clock or oscillator is used to convert the incoming microwave signals to video. The video signals are carried by cable directly to the cross-correlator and the fringes are formed in "real time". Both techniques suffer most from the contribution of the neutral atmosphere to the difference in the signal paths from the different sources.

The most accurate deflection experiment thus far conducted using very-long-baseline interferometry yielded verification of the deflection predicted by general relativity to within the estimated uncertainty of 3% (Counselmann et al., 1974). A connected-element interferometer was used somewhat later to obtain higher accuracy (Fomalont and Sramek, 1976). In this latter experiment, three sources were observed that lie nearly on a line segment 10° in extent on the plane of the sky, with one source being nearly in the ecliptic and nearly half way between the other two. Observations were made at two frequency bands, 2.7 and 8.1 GHz, from the ends of a 35 km long baseline. The central source was observed half of the time and the other two sources were alternately observed for the other half. Data were obtained on a total of 12 days during the month surrounding superior conjuction. The analysis of the data yielded $(1 + \gamma)/2 = 1.015 \pm 0.011$.

The post-post-Newtonian (ppN) contribution to the gravitational bending is predicted to be $10''.9 \times 10^{-6}$ (Epstein and Shapiro, 1980) for a signal grazing the limb of the Sun, as compared to 1''.75 for the post-Newtonian (pN) contribution. The principal problems that have been encountered in the experiments to detect the pN contribution have been the effects of the Earth's atmosphere and the solar corona. To avoid the latter, one can utilize extremely high radio frequencies or, better, optical frequencies. To avoid the former, observations should be made above the atmosphere. Thus, we have proposed a satellite with "crossed" optical interferometers (Reasenberg, 1980). Each interferometer would have a pair of telescopes to gather light and bring it together into an instrument that would form and detect fringes. The fringes formed in a beam splitter would be focused and dispersed to produce channelled spectra, an old idea developed by Fizeau and Foucault. From the pattern of the channelled spectra, one could determine the differential path length along the two arms of the interferometer and therefore the angular offset between the interferometer axis and the star direction.

To do a useful experiment, one could use a pair of these interferometers, about 90° apart. Laser metrology would not only monitor the angle between the two interferometers, but would also monitor each optical component, with an apparently achievable error budget of about 0.1 Å over the proposed 10-m length of the instrument. We are examining this instrument from an engineering standpoint, and have so far found no technological problem that would prevent our achieving an accuracy sufficient to detect the ppN contribution to the deflection, but this assessment may simply reflect the early stage of our investigation. We are continuing the engineering analysis, but the project is clearly long-term and we foresee little chance of obtaining results much before the turn of the century.

3.4. Signal Retardation

The "anomalous" time delay, $\Delta\tau$, attributable to the gravitational field of the Sun is given by

$$\Delta\tau = \frac{2r_0}{c}\,\frac{(1+\gamma)}{2}\ln\left(\frac{r_e + r_p + R}{r_e + r_p - R}\right),$$

where r_0 ($\equiv 2GM_\odot/c^2$) is the gravitational radius of the Sun; and r_e, r_p, and R are, respectively, the distances from the Sun to the Earth, from the Sun to the target planet, and from the Earth to the target planet. The coefficient $(1+\gamma)/2$ is unity in general relativity. When the raypath passes close to the Sun, the argument of the logarithmic term becomes $4r_e r_p/d^2$, where d is the impact parameter of the ray (Shapiro, 1964). This term gives rise to the characteristically sharp spike which is very important because it makes this effect relatively easy to distinguish from the various perturbations of the orbits of the Earth and target planet.

From radar data obtained through 1972, it was possible to estimate $(1+\gamma)/2$ with an uncertainty of 4%. The main contribution to the uncertainty was the limited signal-to-noise ratio available. The solar corona had a negligible effect since the radio frequency used in the radar measurements was nearly 8 GHz. Analyzing tracking data from Mariners 6 and 7 yielded an uncertainty of 3% (Anderson et al., 1976). The tracking data from the Mariner 9 mission, combined with the radar data, gave us a 2% uncertainty; in all cases the estimated value was in agreement with the prediction of general relativity to within the estimated uncertainties.

The solar corona posed an important limitation in the accuracy of the experiments that utilized the Mariner spacecraft because they were equipped with a transponder operable only at the relatively low radio frequency of 2.3 GHz. This plasma problem was greatly alleviated, although not solved, in the Viking relativity experiment. With two launches, each yielding an orbiter and a lander, there were four Vikings to observe. Each orbiter had a transponder operable at 2.3 GHz and, in addition, the ability to transmit an 8.4 GHz signal coherent with the 2.3 GHz signal. The landers, firmly emplanted on the surface of Mars, solved the usual spacecraft problem of unmodelled accelerations. But the landers were equipped with transponders that only transmit 2.3 GHz signals. Since the effect of the plasma of the solar corona has been seen to double in only a few hours, it is clear that no deterministic model will be particularly useful for analyzing the time-delay data from the lander for this relativity experiment.

When Mars was near superior conjunction during the Viking Mission, we therefore tried to obtain simultaneous, or nearly simultaneous, observations of both a lander and an orbiter. A 2.1 GHz signal was sent from the ground to a lander and transponded back to the Earth at 2.3 GHz while another signal, sent from a separate tracking station to an orbiter, was transponded at both 2.3 and 8.4 GHz coherently, and sent back to the Earth. For the orbiter "downlink" path, therefore, we had a measure of the integrated electron density. What we wanted, however, was the integrated electron density along the lander uplink *and* downlink paths. Since the downlink paths were not very far apart in space and in time, we could take the information gained from the orbiter downlink and apply it directly to the lander downlink. But even for simultaneous observations of lander and orbiter, the corresponding assumption is not valid for the lander uplink path. For this latter path, the signal passes the Sun at a significantly earlier time. What we did, therefore, was to apply the "thin-screen" approximation. In this approximation, we assume the entire plasma contribution occurs just as the signal passes a screen that contains the Sun and is normal to the Earth-Mars line. Then, aside from the important time difference, the plasma contribution for the lander uplink is the same as the plasma contribution for the orbiter downlink. With a time delay of one Sun-Mars-Sun light-propagation time, we applied the plasma effect deduced from the orbiter data as one of the corrections to the echo delays measured for the Earth-lander path.

Using these plasma corrections, we were able to obtain residuals for the measurements of echo delays to the lander that were moderately free from systematic trends. However, there *are* systematic trends remaining and they represent a serious problem. It is not clear how much we can improve upon this plasma calibration, although we have a few more techniques in mind to try.

Despite its limitations, our analysis has yielded an excellent estimate of γ: $(1 + \gamma)/2 = 1.000 \pm 0.001$. The model used in this analysis contained 24 parameters; aside from γ, we estimated six initial conditions for the Earth and Mars; six coordinates representing the positions of the two landers on Mars; three parameters describing the rotation of Mars and one the rotation of the Earth; and the Earth-Moon mass ratio. The data used included measurements of delays to the landers for the first 400 days following arrival of Lander 1 on Mars. However, in addition to this analysis, we made and studied over 100 other least-squares solutions. Various other parameters were included; the data set was cut in various ways — ends deleted, pieces in the middle deleted, etc. Other types of numerical experiments were also carried out in an attempt to uncover hidden biases. The uncertainty of 0.001 represents the "equivalent" standard error inferred from the results of these numerical experiments.

What is the likelihood that better results will come from further analysis of the Viking data? For several reasons, we think the likelihood is high. First, we now have data obtained during a period of over 1400 days, more than three times as long as the span discussed above. (Although data continue to be gathered, the data-taking rate is now not nearly so high as it was early in the mission.) Second, data were obtained during a second superior conjunction, although those data are not nearly so numerous as for the first conjunction. Third, we intend to include relevant Doppler-tracking data, radar data, and other spacecraft data. Fourth and finally, our recently completed study of plasma-calibration techniques has shown that it is possible to improve the plasma calibration twofold. If our optimism proves justified, our final uncertainty in $(1 + \gamma)/2$ will be about 2 parts in 10^4.

3.5. Perihelion Advance and the Sun's Quadrupole Moment

The measured advance of the perihelion of Mercury was the first test of the general theory of relativity. The contribution of the Sun to this secular advance is given by

$$\delta\varphi = \frac{3\pi r_0}{p}\left[\frac{2 + 2\gamma - \beta}{3} + \frac{J_2 R^2}{r_0 p}\right] \text{rad/rev},$$

where $(2 + 2\gamma - \beta)/3$ is a combination of PPN parameters, identically unity in general relativity; J_2 the (dimensionless) coefficient of the second zonal harmonic of the Sun's gravitational field ("quadrupole moment"); R and r_0 the physical and gravitational radii of the Sun, respectively; and p the semilatus rectum of the orbit. The quadrupole moment of the Sun, because it has not been measured, makes a most troublesome addition to the secular advance.

Among the planets and presently charted asteroids, Mercury is the best to observe to detect the relativistic contribution to the secular advance of the perihelion. In deciding which bodies are best to observe, several factors need to be considered. First is the rate of the advance; Mercury far exceeds the other bodies in that respect. Second is the observability of the advance: a completely circular orbit would be inappropriate; the larger the orbit and the larger the eccentricity, the larger the "handle" that one has to observe. The product of semi-major axis, eccentricity, and rate of advance thus forms a reasonable figure of merit for measurement. This figure is given for each of the inner planets by Reasenberg (1980). Mars appears as the most useful target, after Mercury, although the orbit of Mars is complicated by its close proximity to the asteroid belt.

Whereas there is only one principal target, the contribution of the Sun to the secular advance contains the parameter J_2 and the PPN parameters, γ and β (as well as others, if we consider possible violations of conservation laws and possible preferred-location effects: Will, 1973). Thus, our ignorance of J_2 is the outstanding serious problem that prevents our isolating the relativistic contribution to the advance.

Inference of J_2 from measurements of the visual oblateness of the Sun is difficult; this method has been tried, but the results are in dispute (Dicke and Goldenberg, 1974; Hill and Stebbins, 1975). Inference of J_2 from comparison of results from Mercury and Mars is also difficult; the effect for Mars is very small, and the influences of the asteroid belt on the orbit of Mars makes the interpretation of a measured advance difficult. A third approach that may be useful is to attempt to detect the difference in the predicted effects from relativity and from J_2 on the *periodic* terms in the orbit of Mercury. This approach, too, is difficult. The best approach, if money were no object, would be to track a spacecraft that passes close to the Sun. We will discuss this option below.

The most useful data now available for the determination of the secular advance are the radar measurements of echo delays from Mercury; similar radar echoes from the other inner planets are also useful, especially in the refinement of the Earth's orbit which is of importance for the determination of Mercury's.

From radar observations of the inner planets made through the early 1970's, the combination $(2 + 2\gamma - \beta)/3$ of PPN parameters was estimated to be 1.003 ± 0.005 (mentioned in Shapiro et al., 1976) on the assumption that the contribution of J_2 was negligible. On the other hand, if we attempt to estimate J_2 simultaneously, we find the estimates very highly correlated and thus very susceptible to systematic errors. We have therefore not yet extracted a useful estimate of J_2 from this approach.

The determination of J_2 for the Sun can be accomplished most accurately by placing a spacecraft close to the Sun, as already mentioned. The U.S. National Aeronautics and Space Administration "Starprobe" Mission, now under preliminary consideration, is being designed to approach close to the Sun. In one possible version of such a mission, the spacecraft would be sent from Earth to pass by Jupiter to obtain a "gravity assist". After the Jupiter encounter, the spacecraft would move toward the Sun in a very eccentric orbit with a perihelion distance of four solar radii. The orbital plane would be *perpendicular* to the ecliptic. After several years of flight, the spacecraft would pass by the Sun in less than a day; yet some very interesting results can be obtained from that brief encounter. Foremost, from our point of view, would be the possibility to estimate J_2. Direct improvement in the estimate of $(2 + 2\gamma - \beta)/3$ would be slight. If a clock were on the spacecraft, there would also be the potential for a very good redshift experiment, as discussed earlier.

There are, of course, problems in tracking a spacecraft close to the Sun. The blackbody equilibrium temperature at perihelion will be ~ 2400 K, well above the melting point of steel. The solar wind and solar radiation pressure become very serious as the spacecraft approaches the Sun. In fact, if those problems are not countered, the Starprobe Mission becomes useless for determining J_2. Fortunately, there is a technology, at least partially developed, to allow "drag-free" motion of the spacecraft, i.e. motion free from substantial perturbations by nongravitational forces: A small homogeneous ball is shielded inside the spacecraft so that it can follow a gravitational trajectory. Capacitive probes, or other sensors, and an electronics package allow the position of the ball to be determined and signals to be sent to the thrusters which keep the spacecraft "centered" on the ball. When the latter moves off center, instead of pushing it back, the spacecraft pushes itself to follow the ball. The spacecraft in essence is a small ball with a big slave shield following it.

This technology is very difficult to implement for use in Starprobe. If an acceptable degree of drag-free motion can be obtained, then a sensitivity study we have done becomes applicable (Reasenberg and Shapiro, 1978; see also Anderson et al., 1978). For a particular orbit and a particular set of measurements and measurement errors, we find that the uncertainty in the estimate of J_2 drops off drastically to a few times 10^{-9}, shortly after periapsis passage. This uncertainty is sufficiently small to virtually eliminate the effect of J_2 on the accuracy of our solar-system tests of general relativity.

3.6. Possible Temporal and Spatial Variations of the Gravitational Constant

Dirac (1938), building on the speculations of Milne and Eddington, has discussed extensively a numerical relation that he refers to as the Large Numbers Hypothesis. There are only very few large dimensionless numbers that can be formed from known physical constants. One of these N_1, say, is the ratio of the electric to the gravitational attraction between an electron and a proton; another, N_2, is the age of the universe expressed in units of atomic time:

$$N_1 \equiv \frac{e^2}{G m_e m_p} \simeq 7 \times 10^{39},$$

$$N_2 \equiv \frac{T}{T_A} \simeq 4 \times 10^3 .$$

In these equations, e is the charge of an electron, m_e and m_p are the masses of an electron and a proton, respectively, T is the age of the universe (of the order of the inverse of Hubble's constant, $H_0 \sim 60$ km/sMpc), and T_A $(= e^2/m_e c^3)$ is the unit of atomic time. A third large number, N_3, is the number of baryons in the observable universe, about $(10^{39})^2$.

Since such large numbers are unusual in physics, perhaps they are related in a fundamental way:

$$N_1 = k N_2 = k' \sqrt{N_3},$$

where k and k' are constants of order unity. Certainly the age of the universe is not a constant. If we assume that the two similar large numbers remain in a fixed ratio, then one of the other "constants" must be variable. The other variable is often taken to be G (see Dyson, 1972); solar-system dynamics should disclose any evidence for $\dot{G} \neq 0$.

Since data related to the dynamics of the solar system span a time very short compared with the age of the universe, we may use a linearized expression to represent any variation of G with time:

$$G(t) = G_0 + \dot{G}(t - t_0) .$$

Since the fractional change in G is small during an orbital period of a planet, it follows that

$$\frac{2\dot{G}}{G} \simeq -\frac{2\dot{a}}{a} \simeq -\frac{\dot{P}}{P} \simeq \frac{\dot{n}}{n},$$

where a, P, and n are, respectively, the semimajor axis, orbital period, and mean motion of the planet. If we compare this planetary gravitational clock to an atomic clock, and if G is varying with time, then we should see a change with time of the relative rates of the two clocks. If we were to observe a planet from the Earth with radar, and were able to isolate the contribution to the echo delay of the signals attributable to any variation of the gravitational constant, we would find that contribution to vary periodically with an amplitude that grows quadratically with time and that is proportional to $\dot{G}$.

One approach that has been taken to seek evidence for a time variation of G is based primarily on the use of lunar observations. In particular, Van Flandern (1975) has utilized precise timings with an atomic clock of stellar occulations of the Moon to determine changes in the Moon's orbital period or mean motion. Because this period is affected significantly by tidal interactions with the Earth, he also employed classical observations of the Moon and the planets to determine the Moon's orbital period with respect to the gravitational clock provided by the planets. By comparing these two periods, he in effect compared atomic time to gravitational time, using the Moon as an interpolation oscillator. Although conceptually correct, there are many problems with this procedure. The lunar orbit is particularly complicated. In addition, the old classical observations are difficult to interpret reliably.

Van Flandern has published a number of results, based on repeated analyses of virtually the same data. Over a span of a half decade, his estimates of $\dot{G}/G$ have varied from $(-12 \pm 3) \times 10^{-11}\,\mathrm{yr}^{-1}$ through 0 to $(+3 \pm 1) \times 10^{-11}\,\mathrm{yr}^{-1}$ (Van Flandern, 1974 and 1979). Based on our experience with similar types of observations, we believe that it is not possible to obtain a reliable value for a standard error nearly so small as those given by Van Flandern. Further, there is little hope for substantial improvement with the use of the classical observations, these were gathered over a period of a few hundred years and provide a limiting uncertainty which cannot be improved upon within a time scale short compared to a century.

An alternative approach which was, in fact, the first to be used to estimate $\dot{G}/G$ with useful accuracy, is based on radar data which determine planetary positions on an atomic time scale. If $\dot{G} \neq 0$, the orbital phase of a planet would develop an offset which would grow quadratically on an atomic time scale. The corresponding signature in the radar observations would be periodic with an envelope that would increase quadratically with time. Our analysis (Reasenberg and Shapiro, 1976) provided the bound $|\dot{G}/G| \leqq 1.5 \times 10^{-10}$ per year. Preliminary sensitivity studies combining radar and Viking data show that the uncertainty can be reduced to about 10^{-11} per year. Within a year we hope to obtain an estimate of $\dot{G}/G$ with an uncertainty at that level.

We now consider possible spatial variations of the gravitational constant. Long (1974) has correctly noted that there is little experimental basis for the assumption that the value of G applicable in the laboratory is also applicable for astronomical distance scales. He notes that on theoretical grounds it has been argued that G could vary with position over distances small compared to those distances encountered in an astronomical setting.

Long has analyzed 19th century data from terrestrial laboratories. When he plots the estimated values of G *versus* ln (R), where R is the separation between the test masses, he finds that the values of G fall near a straight line with a slope that, at least formally, is significantly different from zero. Unfortunately, as with much old data, there are difficulties in determining reliable values for the uncertainties. The data Long analyzed were obtained by different experimenters working in different laboratories under different conditions and may well contain subtle biases which are not discernible from even a careful reading of the original papers.

Long (1976) has also done a laboratory experiment and has obtained a result which he describes as being consistent with the result he obtained from his analysis of 19th century data. However, this result has not been confirmed. Independent and more accurate laboratory experiments (Spero et al., 1980; Newman et al., 1980) to measure the variation of G with distance yielded a result consistent with G not varying for separations from 2 to 5 cm, as expected from ordinary theory. More experiments are being planned to seek evidence for a spatial variation of G over larger distance scales.

3.7. Precession of Gyroscopes

The possibility of using a gyroscope in Earth orbit to test a prediction from general relativity concerning the dragging of inertial frames was suggested independently by Pugh (1959) and by Schiff (1960). A Stanford group has been developing the technology for this test for about 15 years. There are two principal effects to be measured: the first is the so-called motional effect which is related to the Lense-Thirring precession; for a gyroscope in a low polar orbit, the direction of the spin axis is predicted to change by about $0\farcs05$ per year. The second is the geodetic precession, for which general relativity predicts the direction of the spin axis to change by $7''$ per year. Needless to say, this experiment is extremely difficult to carry out successfully.

The Stanford group has developed a design concept in which two pairs of counter-rotating gyroscopes and a proof mass orbit for two years in a drag-free configuration much like the one discussed earlier. The instrumentation would be cooled by liquid helium. Several areas of technology have been advanced in this development and most of the concepts have been tested successfully in the laboratory. The gyroscope experiment appears ready for flight in Earth orbit (Everitt, 1980).

In the past decades, advancing technologies have made possible new and more accurate tests of the theory of gravitation. So far, the experiments have been consistent with the predictions of general relativity. The decades to come will surely bring even more stringent tests based on technologies which perhaps have not yet even been conceived.

Acknowledgement

This work was supported in part by grants from the National Aeronautics and Space Administration, NGS-7556, and the National Science Foundation, PHY 78-07760.

References

Anderson, J. D., P. B. Esposito, W. Martin, C. L. Thornton, and D. O. Muhleman, Astrophys. J., **200**, 221, 1975.

Anderson, J. D., and E. L. Lau, in "A Close-Up of the Sun", M. Neugebauer and R. W. Davies, ed's, JPL Report 78–70, Pasadena, 1978.

Braginsky, V. B., private communication, 1980.

Braginsky, V. B., and V. I. Panov, Zh. Eksp. Teor. Fiz., **61**, 873, 1971.

Cohen, M. R., and I. E. Drabkin, A Source Book in Greek Science (McGraw-Hill, New York, 1948) p. 217.

Counselman, C. C., S. M. Kent, C. A. Knight, I. I. Shapiro, T. A. Clark, H. F. Hinteregger, A. E. E. Rogers, and A. R. Whitney, Phys. Rev. Lett., **33**, 1621, 1974.

Dicke, R. H., and H. M. Goldenberg, Phys. Rev. Lett., **18**, 313, 1967.

Dirac, P. A. M., Proc. R. Soc. Lond., **A 165**, 199, 1938.

Dyson, F. J., in "Aspects of Quantum Theory", A. Salam and E. P. Wigner, ed's (Cambridge University Press, Cambridge, 1972) p. 213.

Epstein, R., and I. I. Shapiro, Phys. Rev. D., in press, 1980.

Everitt, C. W. F., private communication, 1980.

Fomalant, E. B., and R. A. Sramek, Astrophys. J., **199**, 749, 1975.

Hill, H. A., and R. T. Stebbins, Astrophys. J., **200**, 471, 1975.

Long, D. R., Phys. Rev. **D 9**, 850, 1974.

Long, D. R., Nature, **260**, 417, 1976.

Newman, R., J. K. Haskins, J. Pellam, J. Schultz, and R. Spero, Abstracts of Contributed Papers for the 9th International Conference on General Relativity and Gravitation, **2**, 361, 1980.

Nordtvedt, K., Phys. Rev. **169**, 1014, 1968.

Pettengill, G. H., D. B. Campbell, and H. Masursky, Sci. Amer., **243**(2), 54, 1980.

Pugh, G. E., Research Memo. 11, Weapons Systems Evaluaton Group, U.S. Department of Defense, 1959.

Pound, R. V., and J. L. Snider, Phys. Rev. **B 140**, 788, 1965.

Reasenberg, R. D., in "Cosmology and Gravitation", P. G. Bergmann and V. deSabbata, ed's (Plenum, New York, 1980) p. 317.

Reasenberg, R. D., and I. I. Shapiro, in "Atomic Masses and Fundamental Constants, 5", J. H. Sanders and A. H..Wapstra, ed's, (Plenum, New York, 1976), p. 643.

Reasenberg, R. D., and I. I. Shapiro, in " A Close-up of the Sun", M. Neugebauer and R. A. Davies, ed's, JPL Report 78–70, Pasadena, 1978.

Roll, P. G., R. Krotkov, and R. H. Dicke, Ann. Phys., **26**, 442, 1964.

Schiff, L. I., Phys. Rev. Lett. **4**, 215, 1960.

Shapiro, I. I., Science, **157**, 804, 1967.

Shapiro, I. I., C. C. Counselman, and R. W. King, Phys. Rev. Lett., **36**, 555, 1976.

Shapiro, I. I., Phys. Rev. Lett., **13**, 789, 1964.

Spero, R., J. K. Haskins, R. Newman, J. Pellam, and J. Schultz, Phys. Rev. Lett., **44**, 1645, 1980.

Stein, S. R., and J. P. Turneaure, Proceedings I.E.E.E., 1249, Aug. 1975.

Texas Mauritanian Eclipse Team, Astron. J., **81**, 452, 1976.

Van Flandern, T. C., Mon. Not. R. Astr. Soc., **170**, 333, 1975.

Van Flandern, T. C., Bull. Amer. Astron. Soc., **6** (3), 431, 1974.

Vessot, R. F. C., M. W. Levine, E. M. Mattison, E. L. Blomberg, T. E. Hoffman, G. U. Zystron, B. F. Farrell, R. Decher, P. B. Eby, C. R. Baugher, J. W. Watts, D. C. Teuber, and F. D. Wills, Phys. Rev. Lett., **29**, Dec. 1980.

Will, C. M., in "Experimental Gravitation", B. Bertotti, ed., (Academic Press, New York, 1973), p. 1.

Williams, J. G., R. H. Dicke, P. L. Bender, C. O. Alley, W. E. Carter, D. G. Currie, D. H. Eckhardt, J. E. Faller, W. M. Kaula, J. D. Mulholland, H. H. Plotkin, S. K. Poultney, P. J. Shelus, E. C. Silverberg, W. S. Sinclair, M. A. Slade, and D. T. Wilkinson, Phys. Rev. Lett., **36**, 551, 1976.

Address: Prof. Dr. R. D. Reasenberg and Prof. Dr. I. I. Shapiro
Massachusetts Institute of Technology
Cambridge, MA 02139
USA

Quantization about Classical Background Metrics

G. W. Gibbons (Cambridge/U.K.)

I have been asked to survey the current status of this field and I shall try to cover what seems to me to be the most interesting developments since the last General Relativity Conference at Waterloo in 1977. In a talk of this length it is impossible (even if desirable) to mention everything that has been written on the subject in that time. To the extent that any selection is inevitably subjective I risk the danger of neglecting material which others (or Time) will see in a quite different light. Let me therefore take this opportunity of apologizing at the outset for any unintentional omissions or misrepresentations.

1. The Subject Defined

To begin with it is perhaps a good idea to ask precisely what *is* being reviewed. Several different attitudes to the subject have recently emerged and these affect considerably the questions asked and the methods used to answer them. These attitudes may be summarized under the following headings:

1. Pure External Field Theory
2. $G_{\mu\nu} = 8\pi\langle\hat{T}_{\mu\nu}\rangle$
3. The Background Field Method in Lorentzian Spacetime
4. The "Euclidean" Functional Integral Approach

Let us take these in turn.

1.1. Pure External Field Theory

By this I mean the view that the metric $g_{\mu\nu}$ really is fixed or given to one ahead of time, only "matter" fields are quantized. Although this is, on the face of it, mathematically a completely consistent theory per se no-one would, I imagine, regard it as physically satisfactory since no possibility is allowed for the metric to react to the effect of the matter present. At best it is usually thought of as an approximation to some more exact theory. Recently Duff [1.1] has questioned this extent to which it can ever be regarded as a *consistent* approximation to any more extended theory. I shall comment further on this below.

1.2. $G_{\mu\nu} = 8\pi\langle\hat{T}_{\mu\nu}\rangle$

By this I mean the view that the metric should not be quantized at all and that back reaction effects are computed using regularized expectation values of the stress tensor in whatever state is of interest.

Recently Kibble [1.2] and later Kibble, Randjbar-Daemi and Kay [1.3] have argued that *mathematically* this is a well defined theory though it is not Quantum Mechanics. Indeed Kibble [1.2] points out that it can be regarded as providing a sort of non-linear generalization of Quantum Mechanics in which the Principle of Superposition is abandoned. Physically on the other hand, this theory leads one to a number of difficulties of interpretation and it is by no means obvious that these can be solved. How, for instance, does the spacetime geometry change when a quantum mechanical "measurement" is made? By the full amount corresponding to the actual measurement or by a reduced amount corresponding to its expectation? How does one make sense of the "delayed choice" experiments, so beautifully described by Wheeler at this Conference, in this picture? Furthermore, it would seem possible to violate the Uncertainty Principle for ordinary matter if one does not allow the gravitational field to respond to the full change rather than its expectation value — cf. the famous discussion between Einstein and Bohr which hinges precisely on this point [1.4]. Problems like this seem to me to make this theory physically untenable. On the other hand, "Pure External Field Theory" is presumably a consistent mathematical approximation to it. Finally let me remark that if quantum *fluctuations* are small (i.e. the *dispersion* in the quantum state is small) $G_{\mu\nu} = 8\pi\langle\hat{T}_{\mu\nu}\rangle$ seems to be a reasonable and consistent approximation to describe the gravitational effects of very many particles as for instance in a neutron star. Indeed all current work on neutron stars is based precisely on this idea. Such fluctuations are however not small about the vacuum state.

1.3. The Background Field Method in Lorentzian Spacetime [1.5]

This is of course the standard, obvious, view of the subject although this is not very clearly stated in most papers. One expands *both* the matter fields and the gravitational field about some classical background metric and matter field configuration. Schematically:

$$\hat{\varphi} = \varphi_c + \delta\hat{\varphi}_c$$
$$\hat{g}_{\mu\nu} = g_{\mu\nu}{}^c + \delta\hat{g}_{\mu\nu}$$

where for consistency $\{\varphi_c, g_{\mu\nu}^c\}$ solve the classical equations of motion. $\delta\hat{g}_{\mu\nu}$ represents "gravitons" and an important point (for ciblymade by Duff [1.1]) is that the effects of gravitons are comparable with those of "matter" particles as far as the production of particles and their contribution to closed loops is concerned. That is at this level gravitons cannot be ignored.

A good illustration of the point is provided by Page's computations [1.6] of the graviton emission from black holes where it amounts to 2% of the total for a hole whose mass exceeds 10^{15} g.

Once the basic point has been realized it becomes difficult, if not impossible, to maintain a clear cut distinction between matter and metric fluctuations. They are of the same magnitude and in the quantum theory one expects to be able to make arbitrary field redefinitions which leave physical quantities like the S-matrix or other global objects invariant but which mix the local metric with the matter fields. Furthermore, as has been emphasized by 'tHooft [1.7] typical quantum corrections will involve effective actions in which the curvature tensor couples to the matter field derivatives in such a way as to alter the characteristics of the matter field equations. That is $g^c_{\mu\nu}$ no longer gives the light cone. A striking example of this occurs in the work of Drummond and Hathrell [1.8] who compute one loop vertex corrections to QED in an external gravitational field and find to order e^2/m^2 the effective action

$$\frac{1}{4} F_{\alpha\beta}F^{\alpha\beta} - \frac{e^2}{m^2}\frac{1}{2880\pi^2}\{-5RF_{\alpha\beta}F^{\alpha\beta} + 26R_{\alpha\beta}F^{\alpha\sigma}F^{\beta}_{\sigma} - 2R_{\alpha\beta\mu\nu}F^{\alpha\beta}F^{\mu\nu} - 24F^{\alpha\beta}{}_{;\beta}F_{\alpha\sigma}{}^{i\sigma}\}$$

The term $R_{\alpha\beta\mu\nu}F^{\alpha\beta}F^{\mu\nu}$ gives rise to characteristics which lie outside the light cone defined by $g^c_{\mu\nu}$. This paper raises but does not entirely resolve the difficult question of what precisely is the background metric and how well defined is it in the presence of quantum effects. It would be interesting to see how the characteristics change under a field redefinition of the background metric and electromagnetic field for example.

1.4. Expanding about Riemannian Metrics

In flat space quantum field theory it is customary to evaluate functional integrals by "Wick rotating" to imaginary time and dealing with Euclidean quantum fields. Physical amplitudes are obtained by analytic continuation back to Minkowski space. The analogue in General Relativity is field theory in spaces with a positive definite, Riemannian, metric. The idea that the path integral for gravity is to be evaluated by summing over positive definite metrics is referred to as "Euclidean Quantum Gravity" and leads naturally to a consideration of quantum fluctuations about classical solutions of the Einstein equations ("Gravitational Instantons"). There has been a considerable amount of activity in the field since Waterloo and this will be discussed later. However, now is perhaps the time to make what is an important general point. The realization by particle physicists of the existence of solitons and magnetic mono-poles in Minkowski spacetime and of Yang-Mills Instantons in Euclidean space has led to an enormous interest in expanding the full quantum equations about classical backgrounds. Structurally, Yang-Mills theory and General Relativity have much in common and a great deal can be gained by studying them together or by trying to extend ideas found useful in Yang-Mills theory to Gravitation. Examples of this will appear later.

2. Experimental Aspects

This section is by far the shortest and easiest to write — there are very few experiments. However it is perhaps worth mentioning a few relevant experiments and observations on the behaviour of elementary particles in gravitational fields. Good remarked several years ago [2.1] that the K_L, K_S mass difference is so small that one could rule out the possibility that particles feel the earth's gravity differently from antiparticles. Measurements on non-relativistic neutrons falling under gravity have shown that they do indeed satisfy Schroedinger's Equation [2.2] with spinorial wavefunctions [2.3]. Perhaps more relevant for our present interest is the apparent observations of electron-positron pairs production by classical electromagnetic fields [2.4].

Of course there are numerous astrophysical consequences of the Hawking evaporation by Black Holes including the interesting remark by Turner [2.5] that interactions in Grand Unified Theories may cause the asymmetric production of baryons over antibaryons. This offers some exciting prospects for cosmology [2.6] but is a little outside our present topic.

3. Particle Creation

This is usually thought of as *pair production*. However it has become clear that under suitable circumstances this description may not be correct. Labonte [3.1] and later Wald [3.2] realized that the general theory of fermion Bogoliubov transformations allowed the possibility of "strong Bogoliubov transformations" which would correspond to the creation of particles and antiparticles (not necessarily in equal numbers) with absolute *certainty* rather than in pairs with a certain *probability* as is usually the case. Unaware of this work I found an explicit example of this phenomenon [3.3] and was able to relate this to the gravitational anomaly in the conservation of the axial current [3.4]. The parallel phenomenon in Yang-Mills theory was discovered independently by Christ [3.5]. Later work by myself and Richer [3.6] has shown that gravitationally single fermions can be created — a phenomenon which can occur in Yang-Mills theory.

Adopting an abbreviated notation we expand the field operator $\hat{\psi}$ as

$$\hat{\psi} = p^{\text{in}}\hat{a}_{\text{in}} + n^{\text{in}}\hat{b}^{+}_{\text{in}} \tag{3.1}$$

$$= p^{\text{out}}\hat{a}_{\text{out}} + n^{\text{out}}\hat{b}^{+}_{\text{out}} \tag{3.2}$$

where p^{in} and p^{out} are positive frequency solutions in the past or future respectively and n^{in} and n^{out} are negative frequency solutions. $\hat{a}$ and $\hat{b}$ are the corresponding annihilation operators.

$$(p^{\text{out}}, n^{\text{out}}) = (p^{\text{in}}, n^{\text{in}}) \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{3.3}$$

$$= (p^{\text{in}}, n^{\text{in}})\, S \tag{3.4}$$

where S is the classical S-matrix. Then

$$\begin{pmatrix}\hat{a}_{\text{in}}\\ \hat{b}^{+}_{\text{in}}\end{pmatrix} = \begin{pmatrix}A & B\\ C & D\end{pmatrix}\begin{pmatrix}\hat{a}_{\text{out}}\\ \hat{b}^{+}_{\text{out}}\end{pmatrix} \tag{3.5}$$

In order to preserve the commutation relations and conserved inner products S must satisfy the unitarity constraints

$$S^{+}GS = G \tag{3.6}$$

$$SGS^{+} = G \tag{3.7}$$

where $^{+}$ denotes transposed conjugate and for

BOSONS

$$G = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \tag{3.8}$$

But for

FERMIONS

$$G = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \tag{3.9}$$

These yield the following constraints for the Bogoliubov coefficients

BOSONS

$$AA^{+} - BB^{+} = 1 \tag{3.10}$$

$$CA^{+} - DB^{+} = 0 \tag{3.11}$$

$$CC^{+} - DD^{+} = -1 \tag{3.12}$$

$$A^{+}A - C^{+}C = 1 \tag{3.13}$$

$$B^{+}A - D^{+}C = 0 \tag{3.14}$$

$$B^{+}B - D^{+}D = -1 \tag{3.15}$$

But for

FERMIONS

$$AA^{+} + BB^{+} = 1 \tag{3.16}$$

$$CA^{+} + DB^{+} = 0 \tag{3.17}$$

$$CC^{+} + DD^{+} = 1 \tag{3.18}$$

$$A^{+}A + C^{+}C = 1 \tag{3.19}$$

$$B^{+}A + D^{+}C = 0 \tag{3.20}$$

$$B^{+}B + D^{+}D = 1 \tag{3.21}$$

We introduce the operator $\hat{F}$ by

$$\hat{F} = (\hat{a}^+)^+ M\hat{b}^+ \tag{3.22}$$

where for

FERMIONS

M is a skew symmetric matrix

but for

BOSONS

M is a symmetric matrix

$$[\hat{F}, \hat{a}] = -Mb^+ \tag{3.23}$$

$$[\hat{F}, \hat{b}] = -Ma^+ \tag{3.24}$$

thus

$$e^{-\hat{F}}\hat{a}e^{\hat{F}} = \hat{a} + M\hat{b}^+ \tag{3.25}$$

$$e^{-\hat{F}}\hat{b}e^{\hat{F}} = \hat{b} + M\hat{a}^+ \tag{3.26}$$

defining the state $|B\rangle$ by

$$|B\rangle = \exp(-\hat{F}\,|0_-\rangle), \tag{3.27}$$

where $|0_-\rangle$ is the "in-vacuum" and $\hat{F}$ out is constructed from the out creation operators. The in-vacuum $|0_-\rangle$ satisfies

$$\hat{a}_{in}\,|0_-\rangle = 0 \tag{3.28}$$

$$\hat{b}_{in}\,|0_-\rangle = 0 \tag{3.29}$$

that is

$$\begin{aligned} &\{A\hat{a}_{\text{out}} + (AM + B)\,\hat{b}^+_{\text{out}}\}\,|B\rangle = 0 \\ &\{(\overline{C} + \overline{D}M)\,\hat{a}^+_{\text{out}} + \overline{D}\hat{b}_{\text{out}}\}\,|B\rangle = 0 \end{aligned} \tag{3.30}$$

If det $A \neq 0$ we have a *Weak Bogoliubov Transformation.* We choose

$$M = -A^{-1}B \tag{3.31}$$

and

$$|B\rangle = \exp \mathrm{i}W\,|0_+\rangle \tag{3.32}$$

whence

$$|0_-\rangle = \exp(-\mathrm{i}W)\exp\hat{F}\,|0_+\rangle \tag{3.33}$$

$$|\langle 0_+ \mid 0_-\rangle|^2 = \exp(-2/mW) \tag{3.34}$$

and for

BOSONS

$$|\langle 0_+ \mid 0_- \rangle|^2 = |\det A|^{-2} \tag{3.35}$$

while for

FERMIONS

$$|\langle 0_+ \mid 0_- \rangle|^2 = |\det A|^{+2} \tag{3.36}$$

The weak Bogoliubov transformations correspond to *pair* creation since $|0_-\rangle$ contains out-pairs. For bosons the unitarity constraint (3.10) shows that det A is never zero (it may well be infinite) and so *bosons are always created in pairs*[1]). On the other hand for fermions the unitarity constraint (3.16) allows the possibility that det A should vanish. If this is the case then the Bogoliubov Transformation is said to be *Strong*. If $H_{\text{in}}^{(+)}$ is the one particle in Hilbert space and similarly $H_{\text{in}}^{(-)}$, $H_{\text{out}}^{(+)}$ etc. we have that

$$A\colon\ H_{\text{out}}^{(+)} \to H_{\text{in}}^{(+)}$$

$$B\colon\ H_{\text{out}}^{(-)} \to H_{\text{in}}^{(+)}$$

$$C\colon\ H_{\text{out}}^{(+)} \to H_{\text{in}}^{(-)}$$

$$D\colon\ H_{\text{out}}^{(-)} \to H_{\text{in}}^{(-)}$$

The unitarity equations show that C provides a bijection of the Kernel of A into the Kernel of D, and B provides a bijection of the Kernel D into the Kernel of A. Now restricting equation (3.30) to the complement of the image of A (which is non empty since det $A = 0$) we have

$$B\hat{b}_{\text{out}}^{+}\,|B\rangle = 0 \tag{3.37}$$

which will be satisfied if $|B\rangle$ has all antiparticle states in the Kernel of D occupied. Similarly equation (3.30) requires that all particle states in the Kernel of A be filled. On the image of A we may invert A and construct an $\hat{F}$ which creates pairs. Thus

$$|0_-\rangle = e^{\hat{F}}\,|B\rangle \tag{3.38}$$

and $|0_-\rangle$ contains dim Ker A particles and dim Ker D antiparticles with certainty. In fact

$$\langle 0_+ \mid 0_- \rangle = 0 \tag{3.39}$$

---

[1]) Audretsch [3.15] has given an interesting example of a Robertson Walker universe for which $A = I$ and $B = 0$ for bosons and yet presumably vacuum polarization effects are still present.

The net *excess* of particles over antiparticles is given

$$\Delta N = \dim \operatorname{Ker} A - \dim \operatorname{Ker} D \tag{3.40}$$

$$= \dim \operatorname{Ker} A - \dim \operatorname{Ker} A^{+} \tag{3.41}$$

$$= \operatorname{index} A \tag{3.42}$$

where (3.42) defines the Index of the Fredholm operator A. This discussion is essentially that given by Christ [3.5]. I have repeated it here at length because most reviews of the subject cover the boson case and omit this fermion case thus missing some interesting physics. The final result may be easily understood in terms of "hole theory". Kernel A can be thought of as outgoing positive frequency waves which had no positive frequency component in the past and similarly for the Kernel D with positive changed to negative. Since in hole theory such states were filled in the past they must be filled with certainty in the future.

The above decision is general. A specific example in the context of cosmological particle production was given in [3.3], [3.4] and [3.7] for neutrinos where it leads to a violation of the neutrino number. This violation is in fact the well known anomaly in the conservation of the chiral current in a different guise.

The basic idea is that for a metric of the form

$$ds^2 = -dt^2 + g_{ij}(x, t)\, dx^i\, dx^j \tag{3.43}$$

the instantaneous energy eigenvalues and eigenstates will not in general be in 1—1 correspondence if parity is broken. As time varies the energy-eigenvalues will change and may even change sign. Such sign changes or "level crossings" correspond to strong Bogoliubov transformations and the total number with due regard being paid to sign (called by mathematicians the "spectral flow"), gives the total excess of particles over antiparticles created. This turns out to be even. If however one adds cross terms to the metric:

$$ds^2 = -(dt + \omega_i\, dx^i)^2 + g_{ij}(x, t)\, dx^i\, dx^j \tag{3.44}$$

it can be odd [3.7].

Another interesting development in the general area of particle creation is the realization by Birrell and Davies [3.8] that in interacting field theories the well known argument of Parker and Zeldovich [3.9], [3.10] that there should be no creation of non-interacting conformally invariant particles in conformally flat spacetimes, breaks down when one considers the effects of the renormalization of coupling constants. This renormalization introduces a mass scale — call it μ — and thus breaks conformal invariance unless by chance the renormalization group β-function vanishes. The result is that "conformally invariant" particles can be creat edeven in Robertson-Walker spacetimes. Further aspects of the effects of interaction upon particle creation are given by Birrell, Davies and Ford [3.11].

Other aspects of particle creation which have recently received some attention are production by "white holes" by Wald and Sriram Ramaswamy [3.12] and of massive particles in anisotropic cosmologies [3.13]. The whole area of cosmological particle production has been ably reviewed by Hu [3.14] recently and the reader is referred there for further information.

4. Renormalizability of Interacting Field Theory about Fixed Backgrounds

It is well known by now that non-interacting field theory on fixed backgrounds gives rise to divergences in for example the expectation value of the stress tensor $\langle \hat{T}_{\mu\nu} \rangle$ which *cannot* be eliminated by a renormalization of the cosmological constant Λ and Newton's constant G but which can be eliminated if one introduces new terms in the gravitational action proportional to R^2 and $\left(R_{\alpha\beta}R^{\alpha\beta} - \frac{1}{3}R^2\right)$. Recent reviews of the methods used to establish these results are given in [4.1] and [4.2]. The introduction of "boundaries" can change this situation and introduces (at least if one uses standard Dirichlet or Neumann boundary conditions) new infinities. Some of these can be regarded as coming from surface contributions to the gravitational action but others remain — even in flat space [4.3], [4.4].

The question naturally arises whether or not interactions introduce any new infinities. Does a theory which is renormalizable in flat space remain so in curved space? This latter question has been answered (as far as the Green's functions of the theory are concerned) affirmatively by Freedman and Pi [4.5] using perturbation theory about flat space and by Birrell and Taylor [4.6] using generalizations of techniques used in flat space. These results enable one to conclude that using the Thermal Green's function techniques introduced by Perry and myself [4.7], that black holes can remain in equilibrium with a bath of interacting thermal radiation at the black hole temperature. This does *not* mean however that the spectrum of emitted particles is Planckian since particles escaping from the horizon can interact with each other (just as radiation escaping from the centre of a star can) to cause deviations. Some new work by Hawking [4.8] provides a framework for computing this effect.

In addition to these general analyses there exist discussions of the renormalization of vacuum-vacuum amplitudes in Robertson-Walker Universes by Bunch, Panangaden and Parker [4.9], [4.10] for $\lambda\varphi^4$ theory to second order using the concept of normal ordering and adiabatic regularization, and in a general space, again to second order using a momentum space technique, by Bunch and Parker [4.11]. This latter technique which bears a close relation to the Wigner method in quantum mechanics and the theory of pseudo differential operators in Pure Mathematics appears to be a powerful means of analyzing curved space wave equations.

For a more detailed discussion of the renormalization of interacting theories in curved space and other aspects of the interacting field theory in curved space the reader is directed to the recent review by Birrell [4.12].

5. Topological Effects on Lorentzian Spacetimes

There has recently been a considerable upsurge of interest in topological effects. This interest stems from a widespread suspicion that spacetime is not smooth and flat on small scales but may have an extremely complicated topological structure on scales comparable with and perhaps much smaller than the Planck length $\left(\frac{\hbar G}{c^3}\right)^{1/2} \simeq 10^{-33}\,\text{cm}$. This is because (as has been repeatedly emphasized by Wheeler) one expects large quantum fluctuations of the metric at this scale and below. One might also think of these fluctuations — in some loose sense — as "virtual" black holes [5.1] which might be formed because of fluctuations leading to an excess of virtual gravitons in a particular region of space. Such black holes should then evaporate almost as quickly as they have formed via the Hawking evaporation process. The net effect would be a short lived fluctuation or dislocation in the causal structure of spacetime (see fig. 1).

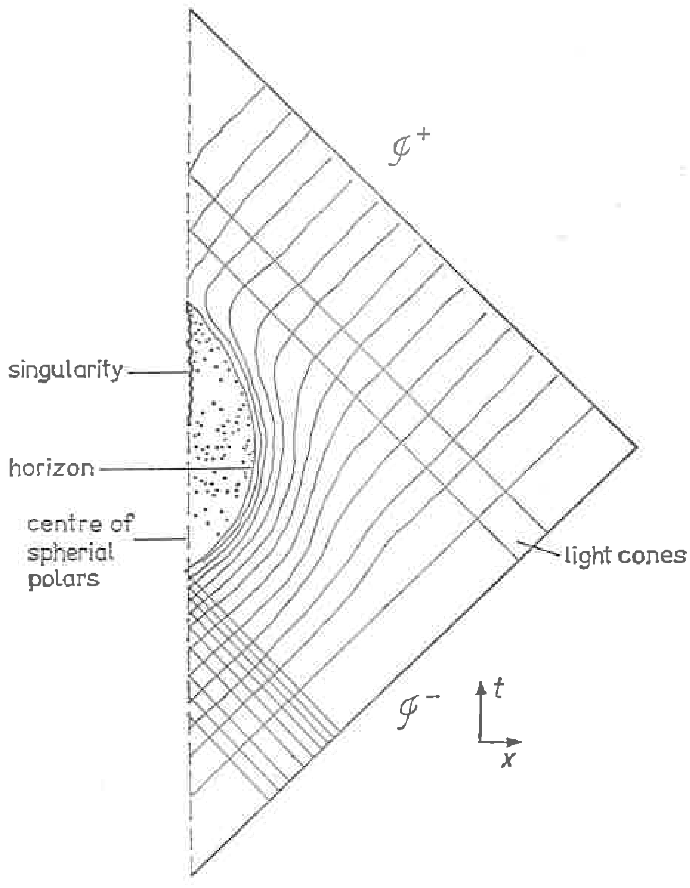


Fig. 1. A virtual Black Hole as a short lived dislocation in the causal structure of space time. N.B. Null rays are only at 45° near the true plot.

If this picture is correct it might help us to overcome the problem of the divergences in quantum field theory and quantum gravity. A foamlike structure at Planck scales and below might produce an effective cutoff on Feynman integrals. These virtual black holes might interact with elementary particles and give rise to observable processes which are otherwise forbidden, for example baryon decay [5.2], [5.3], [5.4]. For instance a virtual black hole might swallow a proton and spit out a positron and some photons. Only conservation laws associated with long range Yang-Mills type fields would still hold in the presence of gravity. Of course baryon decay is now permitted in Grand Unified Theories which put the quarks and leptons in the same representations. One might speculate that these virtual black holes "wormholes" might give rise to other interesting processes [5.5], particularly if they can carry a handedness, analogous to those described in section 3.

The problem with trying to implement some of these ideas in Lorentzian spacetimes are however formidable. To develop Quantum Field Theory convincingly, the spacetime must be globally hyperbolic which implies that its topology is $R \times \Sigma$ where R is the time and Σ some spacelike surface. In particular the topology cannot change [5.6]. The existence of the Lorentzian metric or the existence of a foliation both imply the existence of an everywhere non vanishing vector field. For closed manifolds (compact without boundary) this requires a vanishing Euler number, χ, and in fact the vanishing of χ for a closed manifold is the necessary and sufficient condition for the existence of a foliation of codimension one [5.7]. From this it would appear that the Euler characteristic can play no role in the quantum field theory on globally hyperbolic spacetimes. One way to evade this obstruction is to move into the "Euclidean Regime". Indeed it seems to me a compelling reason for pursuing the Euclidean approach. This will be the subject of later sections, in the meantime I wish to consider what has been done in Lorentzian spacetimes.

There are essentially two sorts of topological effects that can arise in such spaces.

1) Effects arising purely because the topology is non trivial
2) Effects arising because the space in which the fields take their values is non trivial. These effects can only happen if Σ is topologically non trivial.

1) If Σ is not simply connected $\left(\pi_1(\Sigma) \neq 0\right)$ we can move to the universal covering space $\tilde{\Sigma}$. Typically one can regard Σ as an identification space obtained by identifying points on $\tilde{\Sigma}$ under the action of a discrete group of isometries Γ i.e. $\Sigma = \tilde{\Sigma}/\Gamma$. It is then the case that the fundamental group $\pi_1(M)$ is isomorphic to Γ. (An excellent account of the basic mathematical concepts required here is given by Dowker [5.8]). If $\tilde{\Sigma}$ is compact the Poincare Conjecture implies that it is homeomorphic to S^3. Non compact $\tilde{\Sigma}$ can frequently be regarded as being homeomorphic to S^3 with N points removed. For asymptotically flat manifolds, $\tilde{\Sigma}$ the number of asymptotic regions is N. If Σ is compact the homology groups $H_1(\Sigma)$ and $H_2(\Sigma)$ are isomorphic by Poincare Duality and are (by Hurewitz Theorem) given by the abelianization of Γ that is $\Gamma/[\Gamma, \Gamma]$ where $[\Gamma, \Gamma]$ is the commutator subgroup of Γ $(\{g : g = aba^{-1}b^{-1};\ a, b \in \Gamma\})$. $H_2(\Sigma)$ is isomorphic with b_2 copies of Z the group of integers together with copies of cyclic groups (the torsion subgroup). Geometrically b_2, the second Betti number, gives the number of independent 2-spheres which cannot be shrunk to a point. Commonly encountered examples are the "elliptic" spaces where $\tilde{\Sigma} = S^3$ and Γ a suitable subgroup of SO(4). Table 1 shows some typical 3-manifolds and their homotopy and homology groups.

The relation between quantum field theory on spacetime and on its universal covering space has been much studied by Dowker and Banach and Unwin under the name of "Automorphic Field Theory" [5.9] to [5.15]. The basic tool is the "method of images". Let $G(x, y)$ be a Green's function on the universal covering space $\tilde{\Sigma}$ then $G(x, y) = \sum_{g \in \Gamma} G(x, gy)\, a(g)$ is a Green's function on $\Sigma = \tilde{\Sigma}/\Gamma$ where $a(g)$ is some unitary representation of Γ. The introduction of $a(g)$ allows one to introduce "twisted fields". That is we contemplate fields which are not functions but sections of some fibre bundle over spacetime.

Table 1 *Some compact 3 manifolds*

Σ	$\Pi_1(\Sigma)$	$H_1(M, Z)$	b^2	$H_1(M, Z_2)$	
S^3	0	0	0	0	3-sphere
RP^3	Z_2	Z_2	0	Z_2	real projective space
$S^1 \times S^2$	Z	Z	1	Z_2	wormhole
$S^1 \times S^1 \times S^1$	Z + Z + Z	Z + Z + Z	3	$Z_2 + Z_2 + Z_2$	Torus
S(p)	Z^p	Z^p	p	$(Z_2)^p$	p-wormholes
L(p, q)	Z_p	Z_p	0	0	Lens Space $p = 1 \bmod 2$
L(p, q)	Z_p	Z_p	0	Z_2	Lens Space $p = 2 \bmod 4$
L(p, q)	Z_p	Z_p	0	Z_2	Lens Space $p = 0 \bmod 4$
S^2/D^*_n	D^*_n	Z_4	0	Z_2	Binary Dihedral Space $n = 1 \bmod 2$
S^3/D^*_n	D^*_n	$Z_2 + Z_2$	0	$Z_2 + Z_2$	Binary Dihedral Space $n = 2 \bmod 4$
S^3/D^*_n	D^*_n	$Z_2 + Z_2$	0	$Z_2 + Z_2$	Binary Dihedral Space $n = 0 \bmod 4$
S^3/T^*	T^*	Z_3	0	0	Binary Tetrahedral Space
S^3/O^*	O^*	Z_2	0	Z_2	Binary Octahedral Space
S^3/I^*	I^*	0	0	0	Binary Icosahedral Space

2) By now it is generally recognized that mathematically "fields" should be regarded as sections of a suitable fibre bundle E with projection map π and whose base space is spacetime M. The simplest case to consider is when the bundle is trivial. Topological effects then arise from the topology of M because the fibres have non trivial topology. The latter case includes various generalizations of the non-linear-model in which the fields take their values in spheres or other similar spaces [5.16].

Since in 4-dimensions these models are not renormalizable they do not provide very useful models in particle physics and rather more interesting it seems to me are Yang-Mills type theories where the bundle is either a principal or associated bundle with structural group G, and G is a Lie Group. Another possibility is that G is a discrete group like Z_2 and one obtains the twisted fields introduced by Isham [5.17].

The classification of all principal G-bundles over spacetime has been discussed in an important article by Avis and Isham [5.18]. The results have been summarized in [5.19]. One finds that Z_2 bundles are possible only if the manifold is not simply connected and are in 1—1 correspondence with elements[1]) of $H^1(M, Z_2)$. In terms of the auto-

[1]) The number of elements in $H_1(M, Z_2)$ ist called the Moebiusity of M.

morphic fields approach $a(g) = \pm 1$ depending on the path With Lie groups the solution is much richer. If G = SO(n), Sp(n), G_2, F_4 and E_8 the bundles are classified by the familiar instanton number or second Chern number, provided M is compact and orientable, otherwise they are all trivial unless special boundary conditions are imposed. The group U(1) is especially interesting since we are now talking about electromagnetism. As is well known non trivial U(1) bundles are specified by elements of $H^2(M, Z)$. By de Rham theory these may be represented by closed but inexact Maxwell fields whose flux satisfies the Dirac Quantization condition. In Lorentzian spacetime one is in effect talking about quantized versions of the wormholes introduced by Misner and Wheeler in their well known paper on "Geometrodynamics" [5.20].

The quantization of the Maxwell field in spacetimes for which $H_2(M, Z) \neq 0$ presents some interesting new features. A simple example would be the Schwarzschild solution. If one does not quantize the magnetic charge P one has a two parameter family of "vacua". Transitions between the vacua are not allowed because of a "superselection rule" and the representations of the CCR's are not Fock ones. This situation has been analyzed by Ashtekar and Sen [5.21]. Each vacuum state corresponds to a spontaneous breakdown of duality invariance. Sorkin [5.22] has also studied this problem. If one introduces charged matter fields one has no option but to quantize the magnetic monopole moment. One then has the possibility of gravitationally induced CP violating effects [5.23]. Let me here remark parenthetically that in Riemannian metrics there is no distinction between magnetic and electric charge and indeed the electric charges are also quantized [5.24], [5.25].

The zero point energies of twisted scalar fields have been discussed by Hart, Isham and de Witt [5.19] and Banach and Dowker [5.13]. I have tabulated some of the results from this latter reference in Table 2 which nicely illustrates the effects of spacetime topology and twisting. In general it appears that twisting always increases the zero point energy. Note that the sign of the zero point energy can be either positive or negative. For interacting fields more dramatic effects are possible. Spontaneous symmetry breaking in flat space depends upon fields whose values are constant and non vanishing. These correspond to an everywhere non vanishing section of the fibre bundle and will not exist if the bundle is non trivial. Thus in a topologically non trivial background the vacuum state will not be homogeneous. This and the stability of the twisted vacua are discussed in some detail for the case of $\Sigma = S^1 \times R^2$ by Avis and Isham [5.26].

Another quantum effect on scalars which arises in spaces with non trivial topology is the possibility of "topological mass generation" discovered by Ford and Yoshimura [5.27], [5.28] and elaborated upon by Toms [5.29], [5.30]. The basic idea is to start with a theory like $\lambda\varphi^4$ which has only dimensionless coupling constants and consider quantum corrections in a region of size L by evaluating the effective potential for φ. These give rise to terms quadratic in φ with a coefficient of order L^{-2}. They can either be mass terms (if they have one sign) or if they occur with the opposite sign they can cause spontaneous symmetry breakdown. Clearly for cosmological distances (i.e. closed universes) the mass is negligible. It may not have been in the past, cf. [5.31]. At the Planck length it is enormous. This might lead to difficulties if one tried to introduce scalar fields into a theory of spacetime which incorporated the idea of topological structure at Planck scales. Certainly this and the Avis Isham example indicate that the Higgs mechanism in topologically non trivial spacetimes can be very different from what it is in flat spacetime (cf. [6.37], [6.38]).

Table 2 *Scalar zero point energies on some homogeneous 3-manifolds, radius a*

Σ	Moebiosity	Untwisted	Twisted
S^3	1	$\frac{1}{240a}$	/
RP^3	2	$\frac{-7}{240}\,\frac{1}{a}$	$\frac{1}{30}$
S^3/Z_4	2	$\frac{-67}{480}\,\frac{1}{a}$	$\frac{53}{480}\,\frac{1}{a}$
S^3/D^*_2	2	$\frac{-187}{960}$	$\frac{53}{960}$
S^3/T^*	2	$\frac{-3761}{8640}\,\frac{1}{a}$	
$S^3/0^*$	2	$\frac{11321}{17260}$	$\frac{3799}{17280}$
S^3/I^*	/	$\frac{-43553}{43200}$	/

A potentially even more dramatic effect of $\pi_1(M)$ has been discussed by Kiskis [5.43]. Consider action invariant under O(2) — e.g. 2 scalar fields, φ_1 and φ_2. One might wish to define a charge by assigning say charge $+1$ to the combination $\frac{1}{\sqrt{2}}(\varphi_1 + i\varphi_2)$ and -1 to $\frac{1}{\sqrt{2}}(\varphi_1 - i\varphi_2)$. This can always be done locally but if the O(2) bundle is non trivial it may not be possible to do it globally (i.e. reduce the bundle's group from O(2) to SO(2)). Thus "charge" cannot be globally defined and hence not globally conserved. This effect is possible only if $\pi_1(M) \neq 0$. A potentially interesting example from the point of view of $N = 8$ supergravity is a theory whose algebra is that of SO(8) which has a finite automorphism group, $\mathcal{D}$, of order 6. This gives rise to inequivalent discrete bundles labelled by $H_1(M, \mathcal{D})$.

The other sort of boson field occurring in Nature are Yang-Mills fields. Topological effects of these fields have received some study in Lorentzian spacetime from the point of view of black holes [5.32] to [5.37] and also in Riemannian spacetime [5.38] to [5.42] from the point of view of instantons but presumably progress in this field will have to await a better understanding of the flat space theory. Nevertheless it should be possible to clarify the question of what Yang-Mills "hair" and monopolesstructure a black hole can support. At present no complete treatment of this exists in the literature.

The various effects associated with spinorial fields will be discussed later.

6. Riemannization

In modern flat space quantum field theory, both in rigorous "constructive" quantum field theory and in the more heuristic approaches, extensive use is made of the Euclidean ansatz [6.1, 6.2]. Green's function and functional integrals are first evaluated on flat Euclidean 4-space and physical values obtained by analytic continuation in the imaginary time variable — the so called Wick rotation. Provided certain conditions hold — the most important of which is the requirement of Reflection Positivity — the analytically continued Green's functions define a quantum field theory satisfying the Wightman axioms. Indeed using Reflection positivity the Hibert space of quantum mechanics can be constructed entirely in geometric terms in Euclidean space without having to analytically continue back to Minkowski space [6.3, 6.4].

It is natural therefore to attempt to carry these ideas over to quantum field theory in a fixed background metric. This has been done in some cases [6.5, 6.6] and has shed much light on the thermodynamic behaviour of event horizons and their quantum radiance [6.7]. In particular it has allowed the introduction of powerful thermal Green's function techniques to discuss interacting fields around black holes [6.8]. Furthermore it has lead to the development of the elegant and powerful technique of zeta-function regularization to deal with functional integrals in curved space [6.10, 6.11].

However the problem with this technique is that most spacetimes do not admit a Riemannian section. One possible way out of this would be to introduce an arbitrary timelike vector field and consider the class of metrics

$$g_{\alpha\beta}(\lambda) = \lambda V_\alpha V_\beta + g_{\alpha\beta} \tag{6.1}$$

One now considers the theory as a function of λ, $\lambda = -2$ corresponding to a positive definite metric (6.11).

Recently Uhlmann [6.12] (see these proceedings) has shown how to apply the idea of Reflection Positivity to curved space. He requires a Riemannian space $\{M, g_{\alpha\beta}\}$ equipped with a reflection map $\theta : M \to M$ such that

1) θ is an isometry of $g_{\alpha\beta}$
2) M may be decomposed into two disjoint regions $M^\pm$ such that $\theta M^\pm = M^\mp$
3) $(\partial M^+) \cap (\partial M^-)$ is a smooth hypersurface

Using θ Uhlmann shows how to construct the Hilbert space of Quantum Mechanics. For spaces with hypersurface orthogonal Killing vectors (i.e. the analytic continuation of static spacetimes with no horizons) Uhlmann's method gives the standard results. In more general cases it may still be useful where other methods fail. However it clearly cannot be generic. Furthermore it is interesting to note that there is a *topo-*

logical obstruction to the existence of such a θ, the Pontryagin number of the manifold. This is because θ reverses parity. The Pontryagin number is an integral over the manifold of $R_{\alpha\beta\gamma\delta}E^{\alpha\beta\mu\nu}R_{\mu\nu}{}^{\gamma\delta}$ plus possible boundary terms. Since the integrand is odd under θ and the boundary terms vanish if the boundary has an orientation reversing isometry [6.13] the Pontryagin number must vanish. It is easy to construct examples of manifolds with non vanishing Euler number, χ, which admit a θ-map. Of course if one wished to construct a Hamiltonian by defining a time coordinate one would also encounter the Euler class as a topological obstruction.

Within the limits of Quantum Field Theory on a fixed background these difficulties described above limit the general utility of Riemannization techniques although much can be done (see e.g. [6.14]).

Where these techniques really come into their own is in the Riemannian approach to Quantum Gravity [6.15, 6.16, 6.17, 6.18]. In this approach one regards the path integral for gravity as being over all Riemannian (positive definite) metrics of a certain class. The boundary conditions satisfied by the metrics correspond to the freedom to pick an arbitrary matrix element in the conventional Hilbert space approach to Quantum Mechanics. In practice it seems that three sorts of boundary conditions are relevant for Quantum Gravity.

1) Asymptotically Euclidean metrics (A.E.)
2) Asymptotically Flat metrics (A.F.)
3) Compact metrics.

These are relevant for the description of the

1) Vacuum State
2) Grand-Canonical Ensemble
3) Volume-Canonical Ensemble

for Quantum Gravity. The definitions of these metrics are as follows.

1. An asymptotically Euclidean metric is one that inside a compact set K tends to the standard flat metric on $\mathbb{R}^4$
2. An asymptotically Flat metric is one that outside a compact set K tends to the standard flat metric on $R^3 \times S^1$ with the time periodically identified.
3. By compact is also meant with boundary sometimes called "closed".

Stationary points of the gravitational action amongst these classes of metrics satisfy the Einstein condition $R_{\alpha\beta} = \Lambda g_{\alpha\beta}$ and are called "gravitational instantons" [6.19]. In addition to these boundary conditions there exist weakened local forms — the ALE [6.20] and ALF [6.21] boundary conditions whose physical interpretation is at present unclear. Accounts of the properties of the known gravitational instantons are given in [6.22] and [6.23].

The action for gravity is unusual in not being positive definite in the Euclidean Regime [6.24] and this shows up when one quantizes around a background field [6.25]. A special prescription must be made to deal with conformal deformations which can decrease the action by arbitrary amounts. Deformations of the metric which change the conformal structure will in general increase the action. If metrics obey the A.E. boundary conditions such changes will *always* increase the action — by the Positive Action Theorem [6.26, 6.27]. For AF metrics however some changes of the conformal structure can decrease the action [6.28, 6.29]. The Positive Action Theorem also shows that the only A.E. instanton is flat space [6.27]. The Schwarzschild-Kerr family of solutions are A.F. instantons. Since their application to the thermal properties is by now familiar I will turn to the other general case of instantons — the closed ones.

The principle application for closed instantons is to the Hawking's volume canonical approach to spacetime Foam [6.30]. In this approach one sums over all compact metrics with a given 4-volume V. This constraint is enforced using a Lagrange multiplier which turns out to be the familiar cosmological constant Λ. However its interpretation is entirely different from the standard one. Remarkably it turns out that there are many closed solutions of Einstein's equations with negative Λ, whereas there is only one known example with $\Lambda = 0$ (the K3 surface) and just four with $\Lambda > 0$. The examples for $\Lambda < 0$ are known indirectly by an existence proof due to Yau which settles the Calabi-conjecture [6.31, 6.32, 6.33]. Given that the metrics exist it is now necessary to quantize the fluctuations about them. Since the background has $\Lambda \neq 0$ these necessitate a modification of the usual techniques [6.33, 6.34]. Because of an unfortunate error in [6.33] corrected in [6.34] the quantum corrections used in [6.30] were incorrect and the (tentative) conclusions reached there must be modified. The point is that for a general background the most one can calculate at present is the scaling behaviour of amplitudes due to the one-loop divergences of the theory. These cause the amplitude Z to scale as μ^γ where μ is the renormalization mass which is necessary to introduce in order to regularize the functional determinants. For a closed instanton with volume V and Euler number χ, γ given by [6.34] (for Einstein's Theory)

$$\gamma = \frac{106}{45}\,\chi - \frac{87}{60}\,C^2$$

where

$$C^2 = \frac{\Lambda^2}{2\pi^2}\,V$$

This can be either positive or negative depending upon the particular instanton. This unsystematic behaviour is radically changed when we turn to supergravity theories (see later). Interpretations have been suggested for other Riemannian backgrounds but no consensus appears to have been reached as yet [6.35, 6.36].

Recently Hawking, Page and Pope have investigated the influence of a microscopic foam-like structure of spacetime on the propagation of particles through it [6.37, 6.38]. They find that whereas particles with spin are scarecely affected, scalar particles are considerably influenced. Sufficiently so that as to suggest to them that any Higgs scalars found in nature must be composite. Since dynamically in flat space a scalar and a second rank antisymmetric tensor are equivalent it would be interesting to see whether the conclusion held for second rank antisymmetric tensors as well, especially since these descriptions are now known to be inequivalent at the quantum level [6.39].

7. Topology of 4-dimensional Riemannian Manifolds

As I have indicated above if one restricts oneself to Globally Hyperbolic Lorentzian spacetimes the topology is restricted to that of the Cauchy surfaces. The object of interest is then $\pi_1(M)$, the fundamental group. If on the other hand one works with more general Lorentzian spacetimes or moves into the Riemannian regime one encounters all the richness of fourdimensional topology. The topology of 4-manifolds is a vast and difficult area of modern mathematics. In this review I shall merely recall a few of the more important features and refer the reader to the recent review of Mandelbaum [7.1] or the earlier and more concise account of Milnor [7.2] which contains all that is needed for the present purposes.

First of all we note that the problem of classifying all closed 4-manifolds is an unsolvable one [7.3]: This is because any finitely presented group can be realised as the fundamental group of some 4-manifold and there exists no effective algorithm for telling when two such groups are isomorphic.

For this reason attempts to classify closed 4-manifolds are restricted to simply connected 4-manifolds and then only up to homotopy type. From a physical point of view one cannot help wondering about the significance of this undecidability property. Imagaine for instance what it would mean if it were (as it is not) the case that all Feynman graphs were not classifiable. Someone doing perturbation theory could not systematically evaluate graphs and know that he had not counted twice or left something out. Perhaps this undecidability reflects a richness and variety of possibilities in natural phenomena which we are at present unable to appreciate. Or possibly some limitations on our knowledge of spacetime. Whatever the physical significance is no progress has yet been made in understanding it and so we shall turn to simply connected closed manifolds.

Having rid ourselves of $\pi_1(M)$ we are left only with $H_2(M, Z)$. Elements of $H_2(M, Z)$ can (if $\pi_1(M) = 0$) be represented as smoothly embedded orientable 2-surfaces (7.1) but not necessarily as 2-spheres. We then have an integer values symmetric bilinear form I^{ij} on $H_2(M, Z)$ obtained by counting algebraically the number of times the surfaces intersect one another or a slightly displaced version of itself. By Poincare duality the inverse matrix, I_{ij} is also integer valued and given by the cup product on $H^2(M, Z)$. Thus

$$I^{iR} I_{Rj} = \delta^i_j \tag{7.1}$$

and

$$\det I^{iR} = 1 = \det I_{Rj} \tag{7.2}$$

Using Hodge de Rham theory we may represent torsion free elements of $H^2(M, Z)$ by closed 2-forms, $F^i{}_{\mu\nu}$. If $\{C_j\}$ is a basis of $H_2(M, Z)$ dual to $F^i_{\mu\nu}$ i.e. such that

$$\int_{C_j} F^i = \delta^i_j, \tag{7.3}$$

we have

$$\int_M F^i \wedge F^j = I^{ij} \tag{7.4}$$

The signature, τ, is obtained by diagonalizing I_{ij} over the reals. It is thus the number of linearly independent self-dual harmonic 2-forms (b_2^+) minus the number of linearly independent antiself-dual harmonic 2-forms (b_2^-). The Hirzebruch Signature Theorem reads

$$\tau = b_2^+ - b_2^- = \frac{1}{48\pi^2} \int_M R_{\alpha\beta\gamma\delta} {}^*R^{\alpha\beta\gamma\delta} \sqrt{g}\, d^4x$$

where $*$ denotes the Hodge dual on the first two indices. The Euler number χ is given by the alternating sum of the Betti numbers

$$\chi = b_0 - b_1 + b_2^+ + b_2^- - b_3 + b_4 \tag{7.5}$$

$$= 2 - b_1 + b_2^+ + b_2^- \tag{7.6}$$

$$= \frac{1}{32\pi^2} \int R_{\alpha\beta\gamma\delta} {}^*R^{\alpha\beta\gamma\delta *} \sqrt{g}\, d^4x \tag{7.7}$$

by the Gauss-Bonnet theorem.

M has spin structure (see section 8.) if the diagonal elements of I_{ij} are even. I_{ij} is then said to be of even type. In that case M admits a tetrad field which is singular at just one point, p. (M is almost parallelizable.) If one surrounds p with a small sphere one gets a map from $S^3 \to SO(4)$ in an obvious way. Thus we get an element of $\pi_3(SO(4)) = Z \oplus Z$ specified by 2 integers $\frac{1}{2}(2\chi \pm 3\tau)$. These are related to the vacuum states of quantum gravity [6.20, 6.18].

If M has spinor structure the Atiyah-Singer Index Theorem gives the number of normalizable solutions of the Dirac equation counted algebraically according to their chirality as

$$n_+ - n_- = -\frac{\tau}{8} \tag{7.8}$$

If $\pi_1(M) = 0$, τ and $\chi - 2$ gives the signature and rank of I_{ij}. One asks

(1) Does this fix I_{ij} up to changes of basis?
(2) Does this fix M up to homotopy type?

If I_{ij} is not definite $(2\chi \neq |\tau|)$ χ and τ fix I_{ij}. For even I_{ij} it must be what it would be for the connected sum of $\frac{1}{24}\left(\chi - \frac{3}{2}|\tau|\right)$ $S^2 \times S^2$'s and $\frac{|\tau|}{16}$ $K3$'s whereas if I_{ij} is odd it must be what it would be for the connected sum of $\frac{1}{2}\left(\frac{1}{3}\chi + \tau\right)$ CP^2's and $\frac{1}{2}\left(\frac{1}{3}\chi - \tau\right)$ $\overline{CP}^2$'s, where CP^2 is complex projective 2-space, $S^2 \times S^2$ the product of two 2-spheres, $K3$ is Kummer's surface and $\overline{CP^2}$ is CP^2 with the opposite orientation. An optimistic *conjecture* would now be that the manifold must be homotopic (or even diffeomorphic) to such a connected sum. Thus one has a picture of simply connected closed 4-manifolds as being built up of basic building blocks comprised of $S^2 \times S^2$'s, $K3$'s and CP^2's [6.37].

As an example consider algebraic hypersurfaces in CP^3 of degree $R \geqq 4$ [7.4].

$$\tau = -\frac{1}{3} R(R^2 - 4) \tag{7.9}$$

$$\chi = R(R^2 - 4R + 6) \tag{7.10}$$

I_{ij} is even as R is even or odd. Yau's results (6.3) now provide Einstein metrics on these surfaces with $\Lambda < 0$ depending upon $\frac{1}{3}(R + 3)(R + 2)(R + 1)$ real parameters.

Thus we have $(R > 4)$

R even $\quad \frac{(R-2)(R+2)R}{48}$ $K3$'s and $\frac{R}{8}(R-4)^2$ $S^2 \times S^2$'s

R odd $\quad RCP^2$'s and $\frac{1}{3}R(R^2-1)$ $\overline{CP^2}$'s

The case $R = 4$ corresponds to $K3$ but with 58 real parameters. Similarly for the complete intersection of r hypersurfaces of degree $d_i (i = 1, \dots r)$ in CP^r we have:

$$\chi = \frac{1}{2}(A^2 + B)\,C \tag{7.11}$$

$$\tau = \frac{1}{3} BC \tag{7.12}$$

where

$$A = \sum_{i=1}^{i=r} \mathrm{d}_i - (r + 3) \tag{7.13}$$

$$B = \sum_{i=1}^{i=r} \mathrm{d}_i^2 - (r + 3) \tag{7.14}$$

$$C = \prod_{i=1}^{i=r} \mathrm{d}_i \tag{7.15}$$

which can be thought of as

$\frac{c}{48}|B| \quad k3 \;\#\; \frac{c}{8}(A^2 + B - 3|B|) \quad S^2 \times S^2$

or

$\frac{c}{12}(A^2 + 3B) \quad CP^2 \;\#\; \frac{c}{12}(A^2 - B) \quad \overline{CP^2}$

depending upon whether it has spin structure or not.

For non-compact manifolds the preceding theory must be modified — using "relative homology". Some aspects of this are described in [6.22]. In both cases the important point is that the topology of the manifold is given by the intersection matrix, and that this describes how the non-trivial 2-surfaces lie in the manifold. These 2-surfaces can be thought of as Riemannian black holes. Indeed in the Schwarzschild case the horizon analytically continues to that 2-surface (called in [6.22] a "bolt") which represents the non-trivial element of $H_2(M)$ and gives the manifold a non-trivial Euler number of 2.

I would like to thank N. Hitchin for supplying (7.11) to (7.15)

8. Spinors in Curved Spacetime

Perhaps one of the most profound discoveries of this century is that Nature makes use of "2-valued" representations of the rotation group — i.e. of spinors. Not surprisingly the introduction of spinors into curved spaces leads to some interesting new features. The basic questions are of course:

1) Can one introduce spinors (globally) at all?
2) How many inequivalent ways are there of doing so?

In Lorentzian spacetimes the answers were given some time ago by Geroch [8.1]. In Riemannian spacetime it has also been known for some time [8.2]. More recently the question has been re-examined by Isham [8.3]. In general the obstruction to elevating the tangent bundle (with structural group SO(3,1) or SO(4)) to a spin bundle (with structural group SL(2, C) or SU(2) $\times$ SU(2)) is the second Stieffel-Whitney class of the tangent bundle. For Globally Hyperbolic Spacetimes this always vanishes and one is left with inequivalent spinor structures which are in 1-1 correspondence with elements of $H_1(M, Z_2)$. The number of these elements is just the "Moebiusity" we encountered in section 5. These inequivalent spinor structures arise because although the bundles as bundles are trivial the spin connection maps are inequivalent [8.3]. One can split the local tetrad rotations into two classes, "big" and "small". The big rotations cannot be lifted to local SL(2, C) or SU(2) $\times$ SU(2) rotations whereas the small ones can. One finds that the big rotations permute the inequivalent spinor structures amongst themselves. If one does not include *all* the spinor structures on the same footing one does not obtain generating functionals for spinors which are independent of local tetrad rotations [8.4].

The problem of inequivalent spinor structures arises from a non trivial fundamental group. The more serious problem of there being no spinor-structure at all can and does arise in simply connected manifolds and can clearly not be eliminated by passing to a covering space. As mentioned before the mathematical criterion for the existence of a spinor structure is that the intersection matrix I_{ij} has only even entries on the diagonals. CP^2 is an example of an Einstein space for which this is not true.

If the third Stieffel-Whitney class of the tangent bundle vanishes (which happens if $\pi_1(M) = 0$), one may use a spin^c structure. Physically this means that one must consider charged spinor fields in a topologically non trivial background electromagnetic and gravitational field. This possibility was discussed for CP^2 by Hawking and Pope [8.5]. It leads to the conclusion that boson fields have charges which are even multiples of some basic unit whilst all fermions have charges which are an odd multiple of this basic unit — the so called "Charge Statistics Relation".

If the manifold is not simply connected spin^c structures cannot be constructed and one must use more general structures. For instance one can introduce a background SU(2) Yang-Mills field [7.4]. The full range of possibilities has been discussed by Avis and Isham [8.6].

We now turn to the quantized theory. Classically massless Dirac fields on a background admitting spinors possess a conserved chiral current,

$$J_\mu{}^5 = \psi^D \gamma_\mu \gamma_5 \psi \tag{8.1}$$

ψ^D is the Dirac adjoint, because the action is invariant under chiral rotations $\psi \to e^{i\alpha\gamma_5}\psi$. However when this current is regularized using Feynman diagrams or point splitting techniques there arises an anomalous divergence and one finds [8.7, 8.8, 8.9, 9.10]

$$\nabla^\mu J_\mu{}^5 = \frac{1}{192\pi^2} R_{\alpha\beta\gamma\delta} \, {}^*R^{\alpha\beta\gamma\delta} \tag{8.2}$$

as before $*$ denotes Hodge dual on the first pair of indices. The right hand side of this expression is the integrand for the Hirzebrach signature theorem and the Atiyah-Singer Index Theorem. The relation between these facts has been discussed in [6.19, 8.11, 8.12, 8.13]. The discussion is similar to that for the closely related phenomenon in Yang-Mills theory. Consider for simplicity a closed Riemannian manifold. (Various boundary effects are discussed in [8.15, 8.16, 8.17].) The Dirac operator $i\gamma^\alpha\nabla_\alpha$ will have non zero eigen values λ_n which come in γ_5 conjugate pairs. That is if

$$i\gamma^\alpha \nabla_\alpha \psi_n = \lambda_n \psi_n \tag{8.3}$$

$$i\gamma^\alpha \nabla_\alpha (\gamma_5 \psi_n) = -\lambda_n (\gamma_5 \psi_n) \tag{8.4}$$

The zero eigen values do not however come in such pairs. One can split them into n_+ modes $\psi_i^{(+)}$ such that

$$\gamma_5 \psi_i^{(+)} = \psi_i^{(+)} \tag{8.5}$$

and n_- modes $\psi_i^{(-)}$ such that

$$\gamma_5 \psi_i^{(-)} = -\psi_i^{(-)} \tag{8.6}$$

The functional integral for fermions is

$$Z[\tilde\eta^D, \eta] = \int d\tilde\psi^D \, d\psi \, \exp \int (\tilde\psi^D i\gamma^\alpha \nabla_\alpha \psi + \tilde\eta^D \psi + \tilde\psi^D \eta) \sqrt{g} \, d^4x, \tag{8.7}$$

where the twiddled fields are independent of the untwiddled ones. Using Berezin's rules one obtains

$$Z[\tilde\eta^D, \eta] = \Big(\prod_n \lambda_n\Big) \prod_i \int \tilde\psi^D{}_i \eta \sqrt{g} \, d^4x \int (\tilde\eta^D \psi_i) \sqrt{g} \, d^4x \, \exp\Big(-\iint \tilde\eta^D G^\perp \eta\Big) \tag{8.8}$$

where

$$G^{\perp} = \sum_n \frac{\psi_n(x)\, \psi_n^D(g)}{\lambda_n} \tag{8.9}$$

is the inverse of $(i\gamma^\alpha \nabla_\alpha)^{\perp}$, the Dirac operator orthogonal to the zero-modes. The extra zero modes give rise (on differentiating with respect to the spinoral sources $\tilde{\eta}^D$ and η) to helicity changing amplitudes [9.17]. This also means that under a chiral rotation of the sources

$$\eta \to e^{i\alpha\gamma_5}\eta \tag{8.10}$$

$$\tilde{\eta}^D \to \tilde{\eta}^D\, e^{i\alpha\gamma_5} \tag{8.11}$$

the functional Z transforms as

$$Z[\tilde{\eta}^D, \eta] \to \exp 2i\alpha(n_+ - n_-)\, Z[\tilde{\eta}^D, \eta] \tag{8.12}$$

Since formally we could have compensated for the change by transforming

$$\psi \to e^{-i\alpha\gamma_5}\psi \tag{8.13}$$

$$\psi^D \to \tilde{\psi}^D\, e^{-i\alpha\gamma_5}\psi \tag{8.14}$$

in the functional integral it has been argued [8.18] that the "measure" $d\tilde{\psi}^D\, d\psi$ is not chirally invariant but changes as

$$d\tilde{\psi}^D\, d\psi \to \exp 2i\alpha(n_+ - n_-)\, d\tilde{\psi}^D\, d\psi \tag{8.15}$$

which is consistent with the Berezin rules. Now under the transformation $\psi \to \exp(-i\gamma_5)\,\psi$ the action changes as

$$\int i\tilde{\psi}^D \gamma^\alpha \nabla_\alpha \psi \to \int i\tilde{\psi}' \gamma^\alpha \nabla_\alpha \psi - \int \alpha(\nabla^\mu J_\mu{}^5)\sqrt{g}\, d^4x \tag{8.16}$$

using integration by parts. Now

$$n_+ - n_- = -\frac{1}{384\pi^2} \int R_{\alpha\beta\gamma\delta} {}^*R^{\alpha\beta\gamma\delta} \sqrt{g}\, d^4x = -\frac{1}{2} \int \nabla_\mu J_5^\mu \sqrt{g}\, d^4x \tag{8.17}$$

which is consistent (see also [8.19]). A different approach is to construct a current $\tilde{J}_\mu^5$ from $G^{\perp}(x, y)$ using point splitting or zeta function regularization. This yields [8.11, 8.12, 8.13]

$$\nabla^\mu \tilde{J}_\mu^5 = -2 \sum_i \psi_i^D \gamma_5 \psi_i - \frac{1}{192\pi^2} R_{\alpha\beta\gamma\delta} * R^{\alpha\beta\gamma\delta} \tag{8.18}$$

Integration of this expression gives the Atiyah-Singer Index Theorem.

From the physical point of view the important points are that the zero modes give rise to helicity flipping amplitudes and that the difference of positive and negative chirality zero modes is a topological invariant. The relation between this Riemannian discussion and what happens in Lorentzian spacetimes where the zero modes correspond to "level crossings" and may be related to Strong Bogoliubov transformations is discussed in [3.3, 3.4, 3.5 and 3.6].

9. Supergravity and Quantum Field Theory in Curved Spacetime

Although the idea of cancelling the infinitely positive zero point energies of bosons against the infinitely negative zero point energies of fermions goes back at least as far back as Pauli [9.1] who thought that the equality of boson and fermion degrees of freedom was unlikely to hold in practice, it was not until 1976 that a supersymmetric theory of gravity was constructed [9.2, 9.3]. Since that time developments have been rapid and I shall make no attempt to cover them all, but refer to Peter Van Nieuwenhuizen's Lecture for a recent review. I merely wish here to focus on effects arising from quantizing about non trivial backgrounds.

The first point to be made is that by there exists a version of simple supergravity which incorporates a possible (negative) cosmological term [9.4, 9.5, 9.6] and that gauging $O(N)$ supergravity for $N = 4$ inevitably gives rise to a Λ term. It is not yet known whether the $N > 4$ models can be gauged. We see now the importance of the considerable body of QFT work done previously in curved spacetimes with non vanishing Λ, [9.7, 9.8]. It is *inconsistent* to quantize these theories about flat spacetime — the background *must* be non trivial. The background field method was used by Yoneya [9.9] to discuss simple supergravity and has been extended recently [9.10] to include backgrounds with $\Lambda \neq 0$, and the results are summarized in [9.11]. The main point is that the scaling parameter γ has the form $x\chi + yC^2$ where x, y:

$N = 0$	$x = 106/45$	$y = -87/60$
$N = 1$	$41/24$	$-77/72$
$N = 2$	$11/12$	$-13/18$
$N = 3$	0	$-5/12$
$N = 4$	-1	-1
$N = 5$	-2	0
$N = 6$	-3	0
$N = 7$	-5	0
$N = 8$	-5	0

The case $N = 0$ is ordinary gravity. The remarkable feature is that the C^2 contributions vanish for $N > 4$. This is unexpected and indicates that the theory is less divergent than might have been expected. Recently it has been pointed out by Siegel [9.12] that spin zero fields used in [9.10, 9.11] were all scalar fields whereas the dimensional reduction from 11-dimensional simple supergravity [9.13, 9.14, 9.15] suggests the use of (for $N = 8$) 63 scalars, 7 2-forms and 1 3-form. Although classically the scalars and 2-forms are equivalent representations of a spin 0 field and a 3-form has no dynamical degrees of freedom Duff and van Nieuwenhuizen [6.39] have pointed out that they have different quantum properties. In particular because of inequivalent ghost structures the trace anomalies differ. When applied to $N = 8$ we obtain the striking result that $\gamma = 0$ [9.16]. There are no quantum corrections even on non trivial backgrounds. Note that this result applies to the known ungauged theory as well as the as yet unknown and postulated gauged theory.

The results just given apply to any background. If one examines particular backgrounds even more striking things arise. In particular self-dual backgrounds are of special interest. The subject was initiated by Hawking and Pope [9.17] who showed that at the one loop level the divergences due to the non zero eigenvalues in the functional determinants all cancelled. To do this they made use of the existence in these spaces of covariantly constant spinors to relate fields of different spin. The relations may in fact be thought of as global supersymmetry transformations on the system being considered. Because of the limited number of invariants one can construct on a self-dual background it was later pointed out by Christensen, Deser, Duff and Grisaru [9.18] and Kallosh [9.19], that higher loop correction must be finite on these spaces. The fascinating combination of Riemannian Geometry, topology and quantum field theory deployed in these studies has encouraged further activity along these lines [9.19, 9.20, 9.21, 9.22].

An as yet only partially explored area is the connection between the twisted fields of Isham and supersymmetry. It appears that twisted scalar fields really are bosonic — nevertheless there are fascinating hints of a deeper relation between supersymmetry and topology. For instance the twisted scalar fields and the inequivalent spin structures on a manifold are both given by elements of $H_1(M, Z_2)$ and paired in such a way as to permit the extension of supersymmetry to non trivial sectors [9.20].

Conclusion

The subject of quantization about background metrics is evidently a flourishing one. One sign of this is the comparitively small overlap of the material covered here and that covered in my earlier review for an Einstein centennial volume which was written in the autumn of 1977. However the direction of research has changed somewhat from the rather narrow confines of quantizing only the matter fields to quantizing gravity as well. Thus one regards the subject as a tool to elucidate the structure of quantum gravity rather than an end in itself. In the case of supergravity for which a non-vanishing Λ term is required it is an essential tool — quantizing about flat space is inconsistent. I believe if the relation to the full quantum theory of gravity is borne in mind one can expect many exciting developments in the future. If it is not then the subject will turn into a stagnant backwater. In particular, the most urgent problem facing this subject is the renormalizability or possible finiteness of quantum gravity. Elaborate computations based on a divergent theory — for instance, based on trace anomalies — seem to me to be misguided. At present the most promising candidate for a finite theory of gravity is the $N = 8$ extended supergravity model. It is completely finite around curved backgrounds to one loop. It seems an appropriate theory to tackle such questions as the back reaction problem to one loop. At least at that level the results are unambiguous.

References[1])

[1.1] M. J. Duff, "Is Quantum Field Theory on a Fixed Background Consistent?". 2nd Oxford Conference on Quantum Gravity and to be published in the proceedings.
[1.2] T. W. B. Kibble, "Relativistic Models of Non-Linear Quantum Mechanics". Comm. Math. Phys. **64**, 73 (1979).
[1.3] T. W. B. Kibble and S. Randjbar-Daemi, "Non-Linear Coupling of Quantum Theory and Classical Gravity". J. Phys. **A 13**, 141—148 (1980).
T. W. B. Kibble, B. Kay and S. Randjbar-Daemi, "Renormalization of Semi-classical field theories".
[1.4] N. Bohr, "Discussion with Einstein on epistemological problems in atomic physics" in Einstein, Philosopher Scientist ed. P. A. Schilpp, The Library of Living Philosophers, Evanston, Ill. 1949 p. 218.
[1.5] B. de Witt, Phys. Rev. **162**, 1195, 1239 (1967) also "Dynamical theory of Groups and Fields" in: Relativity Groups and Topology Les Houches lectures. Gordon & Breach, N.Y. 1973.
[1.6] D. N. Page, "Particle omission rates from a black hole: Massless particles from an uncharged or rotating hole. Phys. Rev. **D 13**, 198—206 (1976).
[1.7] G. 'tHooft, "Quantum Gravity: A Fundamental Problem and Some Radical Ideas" in M. Levy & D. Deser ed. "Gravitation" Cargèse 1978 Plenum Press N.Y. (1979).
[1.8] I. Drummond and S. Hathrell, "Q.E.D. Vacuum Polarization in a Background Gravitational Field and its Effect on the Velocity of Light. Phys. Rev. **D 22**, 343—354 (1980).

[2.1] M. L. Good, "K^0_2 and Equivalence Principle". Phys. Rev. **121**, 311—313 (1961)
[2.2] R. Colella, A. W. Overhauser & S. A. Werner, "Observation of Gravitationally Induced Quantum Interference". Phys. Rev. Letts. **34**, 1472—1474 (1975).
[2.3] S. A. Werner, R. Colella and A. W. Overhauser, "Observation of the phase shift of a neutron due to precession in a magnetic field". Phys. Rev. Letts. **35**, 1053—1055 (1975).
[2.4] H. Backe et al., Phys. Rev. Letts. **40**, 1443 (1978) "Observation of Positron Creation in Superheavy Ion-Atom collision systems".
[2.5] M. S. Turner, "Baryon Production by Primordial Black Holes". Phys. Letts. **89B**, 155 to 159 (1979).
[2.6] J. D. Barrow, "Primordial Baryon Generation by Black Holes". M.N.R.A.S. **192**, 427 to 439 (1980).

[3.1] G. Labonte, "On the Nature of "Strong" Bogoliubov Transformations for Fermions". Commun. Math. Phys. **36**, 59—72 (1974).
[3.2] R. M. Wald, "Existence of the S-matrix in Quantum Field Theory in Curved Space-Time". Ann. of Phys. (N.Y) **118**, 490—516 (1979).
[3.3] G. W. Gibbons, "Cosmological Fermion-Number non Conservation". Phys. Letts. **84B**, 431—434 (1979).
[3.4] G. W. Gibbons, "Spectral Asymmetry and Quantum Field Theory in Curved Spacetime". Ann. of Phys. (N.Y.) **125**, 98—116 (1980).
[3.5] N. Christ, "Conservation Law Violation at High Energies by Anomalies". Phys. Rev. **D 21**, 1591—1602 (1980).
[3.6] G. W. Gibbons and J. Richer, "Gravitational Creation of Odd Numbers of Fermions". Phys. Letts. **89B**, 338—340 (1980).

[1]) Unfortunately we did not receive information on formula (3.27) and the incomplete quotations (ed.).

[3.7] G. W. Gibbons, "Spectral Asymmetry and Quantum Field Theory in Curved Spacetime" in R. Ruffini ed. "Proc. 2nd Marcel Grossman meeting on Recent Developments in General Relativity", Trieste July 1979, to be published by the North Holland Publishing Company.
[3.8] N. D. Birrell and P. C. W. Davies, "Conformal Symmetry Breaking and Cosmological Particle Creation in $\lambda\varphi^4$ Theory". Phys. Rev. **D 22**, 322—329 (1980).
[3.9] L. Parker.
[3.10] Ya. B. Zeldovich.
[3.11] N. D. Birrell, P. C. W. Davies and L. H. Ford, "Effects of Field Interactions upon Particle Creation in Robertson-Walker Universes". J. Phys. **A 13**, 961—968 (1980).
[3.12] R. M. Wald and Sriram Ramaswamy, "Particle Production by White Holes". Phys. Rev. **D 21**, 2736—2741 (1980).
[3.13] N. D. Birrell, P. C. W. Davies, "Massive Particle Production in Anisotropic Spacetimes". J. Phys. **A 13**, 2109—2120 (1980).
[3.14] B. L. Hu, "Quantum Field Theories and Relativistic Cosmology" in R. Ruffini ed. Proc. 2nd Marcell Grossman meeting on Recent Developments in General Relativity, Trieste July 1979, to be published by the North Holland Publishing Company.
[3.15] J. Audretsch, "Is the Cosmological Particle Creation Adequately described by an Asymptotic IN-OUT calculation?" Phys. Letts. **72 A**, 401—403 (1979).

[4.1] D. Boulware, "Quantum Field Theory Renormalization in Curved Spacetime" in S. Deser & M. Levy Ed. "Gravitation Cargèse 1978" Plenum Press 1979.
[4.2] L. Parker, "Aspects of Quantum Field Theory in Curved Space-time: Effective Action and Energy-Momentum Tensor" in S. Deser & M. Levy ed. "Gravitation Cargèse 1978", Plenum Press (1979).
[4.3] D. Deutsch and P. Candelas, "Boundary Effects in Quantum Field Theory". Phys. Rev. **D 20**, 3063—3080 (1979).
[4.4] G. Kennedy, R. Critchley and J. S. Dowker, "Stress Tensor and Surface Action Renormalization" Ann. Phys. **125**, 346—400 (1980).
[4.5] D. Z. Freedman and S-Y. Pi, "External Gravitational Interactions in Quantum Field Theory". Ann. of Phys. **91**, 442—449 (1975).
[4.6] N. D. Birrell and J. G. Taylor, "Analyses of Interacting Quantum Field Theory in Curved Spacetime". J.M.P. **21**, 1740—1760 (1980).
[4.7] G. W. Gibbons and M. J. Perry, "Black Holes in Thermal Equilibrium". Phys. Rev. Letts. **36**, 985 (1976).
[4.8] S. W. Hawking, "A causal propagation in Quantum Gravity" to appear in Proc. of the 2nd Oxford Conference on Quantum Gravity eds. C. J. Isham, R. Penrose and D. W. Sciama.
[4.9] T. S. Bunch, P. Panangaden and L. Parker, "On renormalization of $\lambda\varphi^4$ field theory in curved space-time: I". J. Phys. **13 A**, 901—918 (1980).
[4.10] T. S. Bunch and P. Panangaden, "On renormalization of $\lambda\varphi^4$ field theory in curved space-time: II". J. Phys. **13 A**, 919—932 (1980).
[4.11] T. S. Bunch and L. Parker, "Feynmann Propagator in curved Spacetime: A momentum-space representation". Phys. Rev. **D 20**, 2499—2510 (1979).
[4.12] N. D. Birrell, "Interacting Quantum Field Theory in Curved Spacetime" to appear in "Quantum Gravity II: a Second Oxford Symposium ed. C. J. Isham, R. Penrose and D. W. Sciama O.U.P.

[5.1] A. Ashtekar and R. Geroch, Rep. Prog. Phys. **37** no. 10 (1974).
[5.2] S. W. Hawking.
[5.3] Ya. B. Zeldovich, "A new type of radioactive decay: gravitational annihilation of baryons". Phys. Letts. **59 A**, 254 (1976).
[5.4] J. D. Barrow & F. J. Tipler, Nature (Lond) **276**, 453 (1978).
[5.5] Ya. B. Zeldovich, "Gravitation, the mass of particles and the Instantons". Phys. Letts. **73B**, 423—424 (1978).

[5.6] R. P. Geroch, "Topology in General Relativity". J. Math. Phys. **9**, 450–465 (1968).
[5.7] W. P. Thurston, "Existence of codimensional-one foliations". Ann. Math. **104**, 249 to 268 (1976).
[5.8] J. S. Dowker, "Selected Topics in Topology and Quantum Field Theory". Austin preprint.
[5.9] R. Banach and J. S. Dowker, "Automorphic Field Theory — Some Mathematical Issues". J. Phys. **A 12**, 2527–2544 (1979).
[5.10] R. Banach, "The Quantum Theory of Free Automorphic Fields". J. Phys. **A 13**, 2179 to 2204 (1980).
[5.11] R. Banach, "Topology and Renormalizability".
[5.12] J. S. Dowker and R. Banach, "Quantum Field Theory or Clifford-Klein spacetimes. The effective Lagrangian and vacuum stress-energy tensor". J. Phys. **A 11**, 2255 (1978).
[5.13] R. Banach and J. S. Dowker, "The vacuum stress-tensor for automorphic fields on some flat spacetimes". J. Phys. **A 12**, 2545–2562 (1979).
[5.14] S. D. Unwin, "Quantized Spin; — field in flat Clifford-Klein spacetimes". J. Phys. **A 13**, 313 (1980).
[5.15] S. D. Unwin, "Thermodynamics in Multiply Connected Spaces". J. Phys. **A 12**, L309 to L315 (1979).
[5.16] E. Gava, R. Jengo, C. Omero and R. Percacci, "Classical solutions for a generally covariant CP Non-Linear model". Nucl. Phys. **B 151**, 457–468 (1979). R. Percacci "Geometrical Aspects of Non-Linear σ-models" Trieste preprint.
[5.17] C. J. Isham, "Twisted Quantum Fields in a Curved Spacetimes". Proc. Roy. Soc. **362A**, 383–404 (1978).
[5.18] S. J. Avis and C. J. Isham, "Quantum Field Theory and Fibre Bundles in a General Space-Time" in S. Deser and M. Levy eds. "Gravitation Cargèse 1978" Plenum Press 1979.
[5.19] B. S. de Witt, C. F. Hart and C. J. Isham, "Topology and Quantum Field Theory". Physica **96A**, 197 (1979).
[5.20] C. W. Misner and J. A. Wheeler, "Classical Physics as Geometry, Gravitation, electromagnetism, unquantized charge and mass properties of empty space". Ann. Phys. **2**, 525–603 (1957).
[5.21] A. Ashtekar and A. Sen, "On the role of spacetime topology in quantum phenomena Superselection of charge and emergence of non trivial vacua". J. Math. Phys. **21**, 526–533 (1980).
[5.22] R. Sorkin, "The Quantum electromagnetic field in multiply connected space". J. Phys. **12A**, 403–421 (1979).
[5.23] S. Deser, M. J. Duff and C. J. Isham, "Gravitationally induced CP effects". Phys. Letts. **93B**, 419–423 (1980).
[5.24] M. J. Duff and J. Madore, "Einstein-Yang-Mills pseudo particles and charge quantization". Phys. Rev. **D18**, 2788–2797 (1978).
[5.25] R. Percacci, "Einstein Maxwell Instantons and Charge Quantization". Nuovo Cimento **52B**, 90–104 (1979).
[5.26] S. J. Avis and C. J. Isham, "Vacuum Solutions for a Twisted Scalar Field". Proc. Roy. Soc. **A 363**, 581 (1978).
[5.27] L. H. Ford and T. Yoshimura, "Mass generation by self interaction in non-minkowskian spacetimes". Phys. Letts. **70A**, 89–91 (1979).
[5.28] L. H. Ford, "Casimir effect of self interacting scalar field". Proc. Roy. Soc. **A 368**, 311 (1979).
[5.29] D. V. Toms, "Casimir effect and topological mass". Phys. Rev. **D 21**, 928–932 (1980).
[5.30] D. V. Toms, "Symmetry breaking and mass generation by space-time topology" Phys. Rev. **D 21**, 2805–2817 (1980).
[5.31] G. W. Gibbons, "Symmetry restoration in the Early Universe" J. Phys. **A11**, 1341 to 1345 (1978).
[5.32] F. A. Bais, R. Russell, "Magnetic-monopole solution of non-Abelian gauge theory in curved spacetime". Phys. Rev. **D 11**, 2692–2695 (1975).

[5.33] M. Y. Wang, Phys. Rev. **D 12**, 3069 (1975).
[5.34] G. W. Gibbons, "Black Hole Dyons need not explode". Phys. Rev. **D 13**, 3530 (1976).
[5.36] M. J. Perry, "Black Holes are Coloured". Phys. Letts. **71B**, 234 (1977).
[5.37] S. L. Adler and R. P. Pearson, "No Hair theorem for the Abelian Higgs and Goldstone models". Phys. Rev. **D 18**, 2798 (1978).
[5.38] J. M. Charap and M. J. Duff, "Gravitational Effects on Yang-Mills Topology". Phys. Letts. **69B**, 445–447 (1977).
[5.39] J. M. Charap and M. J. Duff, "Space-Time Topology and a new class of Yang-Mills instanton". Phys. Letts. **71B**, 219 (1977).
[5.40] C. N. Pope and A. Yuille, "A Yang-Mills Instanton on Taub-NUT Space". Phys. Letts. **78B**, 424–426 (1978).
[5.41] A. Yuille, "Yang-Mills Instantons in Curved Spacetime". Phys. Letts. **81B**, 321 (1979).
[5.42] H. Romer, "Number of Parameters of Instanton solutions on arbitrary four-dimensional compact manifolds". Nuovo Cimento Letts. **21B**, 381–386 (1978).
[5.43] J. Kiskis, "Disconnected Gauge Groups and the Global violation of charge conservation. Phys. Rev. **D17**, 3196–3202 (1978).

[6.1] J. Schwinger.
[6.2] K. Symanzik, "Euclidean Quantum Field Theory" in "Local Quantum Theory". R. Jost ed. Proceedings of the International School of Physics Enrico Fermi XLV Corso. Academic Press (1969).
[6.3] "Constructive Quantum Field Theory" eds. G. Velo and A. S. Wightman, Springer Lecture Notes in Phys. no. 25 (1973).
[6.4] J. Glimm and A. Jaffe, in "New developments in Quantum Field Theory and Statistical Mechanics". Cargèse (1967) eds. M. Levy and P. Mitter, Plenum.
[6.5] J. B. Hartle and S. W. Hawking, "Path-Integral derivation of black hole radiance". Phys. Rev. **D 13**, 2188–2203 (1976).
[6.6] G. W. Gibbons and S. W. Hawking, "Cosmological event horizons Thermodynamics and particle creation". Phys. Rev. **D 15**, 2738–2751 (1977).
[6.7] S. W. Hawking, "Black Hole Explosions?" Nature **248**, 30–31 (1974).
[6.8] G. W. Gibbons and M. J. Perry, "Black Holes and Thermal Green's Functions". Proc. Roy. Soc. **A 358**, 467–494 (1978).
[6.9] S. W. Hawking, "Zeta function regularization of path integrals in curved space-time". Commun. Math. Phys. **56**, 133–148 (1977).
[6.10] G. W. Gibbons, "Functional Integrals in Curved Spacetime" in K. Bleuer, H. Petry and A. Reetz eds., Proceedings of the second Bonn Conference on Differential Geometric Methods in Physics, Springer Lecture notes in Mathematics no. 676.
[6.11] P. Candelas and D. J. Raine, "Feynman Propagator in Curved Spacetime". Phys. Rev. **D 15**, 1494 (1977).
[6.12] A. Uhlmann, "On Quantization in Curved Space Time" in Proceedings of the 1979 Serpukhov International workshop on high energy physics and field theory.
[6.13] M. F. Atiyah, W. K. Patodi and I. M. Singe, "Spectral asymmetry and Riemannian Geometry I, II and III, Math. Proc. Camb. Phil. Soc. **77**, 43–69 (1975), **78**, 405–432 (1975), **79** (1976).
[6.14] R. M. Wald, "On the Euclidean approach to quantum Field theory in curved spacetime. Commun. Math. Phys. **70**, 221 (1980).
[6.15] G. W. Gibbons and S. W. Hawking, "Action Integrals and Partitons Function in Quantum Gravity". Phys. Rev. **D 15**, 2752–2756 (1977).
[6.16] S. W. Hawking, "The path-integral approach to quantum gravity" in S. W. Hawking & W. Israel eds. "General Relativity: An Einstein Centenary Survey", C.U.P. (1979).
[6.17] S. W. Hawking, "Quantum Gravity and Path Integrals". Phys. Rev. **D 18**, 1747–1753 (1978).
[6.18] S. W. Hawking, "Euclidean Quantum Gravity" in S. Deser and M. Levy ed. "Gravitation" Cargèse (1978) Plenum Press.
[6.19] S. W. Hawking, "Gravitational Instantons". Phys. Letts. **60 A**, 81 (1977).

[6.20] G. W. Gibbons and S. W. Hawking, "Gravitational Multi Instantons". Phys. Letts. **75B**, 430 (1978).
[6.21] G. W. Gibbons, C. N. Pope and H. Romer, "Index Theorem Boundary Terms for Gravitational Instantons". Nucl. Phys. **B 157**, 377—386 (1979).
[6.22] G. W. Gibbons and S. W. Hawking, "Classification of Gravitational Instanton Symmetries". Commun. Maths. Phys. **66**, 291—310 (1979).
[6.23] G. W. Gibbons, "Gravitational Instantons: A Survey" in Proc. International Conference on Mathematical Physics, Lausanne 1979. K. Osterwalder ed. Springer Lecture Notes in Physics No. 116.
[6.24] G. W. Gibbons, "The Einstein Action of Riemannian metrics and its relation to Quantum Gravity and Thermodynamics". Phys. Letts. **61A**, 3—8 (1977).
[6.25] G. W. Gibbons, S. W. Hawking and M. J. Perry, "Path Integrals and the Indefiniteness of the Gravitational Action". Nucl. Phys. **B 138**, 141—150 (1978).
[6.26] G. W. Gibbons and C. N. Pope, "The Positive Action Conjecture and Asymptotically Euclidean Metrics in Quantum Gravity". Commun. Math. Phys. **66**, 267—290 (1979).
[6.27] R. Schoen and S. T. Yau, "Proof of the Positive Action Conjecture in Quantum Relativity". Phys. Rev. Letts. **42**, 547—548 (1979).
[6.28] D. N. Page, "The Positive Action Conjecture". Phys. Rev. **D 18**, 2733—2738 (1978).
[6.29] M. J. Perry.
[6.30] S. W. Hawking, "Spacetime Foam". Nucl. Phys. **B 144**, 349—362 (1978).
[6.31] S. T. Yau, "Calabi's conjecture and some new results in algebraic geometry". Proc. Natl. Acad. Sci. **74**, 1798—1799 (1977).
[6.31] S. T. Yau, Commun. Pure and Applied Math. **31**, 379—411 (1978).
[6.32] Premiere class de Chern et Courbure de Ricci: preuve de la conjecture di Calabi seminarie Palalseau 1978 Astenique **88** (1978).
[6.33] G. W. Gibbons and M. J. Perry, "Quantizing Gravitational Instantons". Nucl. Phys. **B 146**, 90—108 (1978).
[6.34] S. M. Christensen and M. J. Duff, "Quantizing Gravity with a Cosmological Constant".
[6.35] M. J. Perry, "T. P. Inversion in Quantum Gravity". Phys. Rev. **D 19**, 1720—1731 (1979).
[6.36] A. Ashtekar-Magnon.
[6.37] S. W. Hawking, D. N. Page and C. N. Pope, "The propagation of particles in spacetime foam". Phys. Letts. **86B**, 175 (1979).
[6.38] S. W. Hawking, D. N. Page and C. N. Pope, "Quantum gravitational bubbles". DAMTP preprint
[6.39] M. J. Duff and P. van Nieuwenhuizen, "Quantum inequivalence of different field representations". Phys. Letts. **94B**, 179—182 (1980).

[7.1] R. Mandelbaum, "Four-Dimensional Topology: An Introduction". Bull. A.M.S. **2**, 1—159 (1980).
[7.2] J. Milnor, "On simply connected 4-manifolds". Symp. Internat. de Top. Alg. (Mexico 1956) 122—128 (1958).
[7.3] A. A. Markov, Proc. Int. Congr. of Mathematicians 1958 pp. 300—306.
[7.4] A. Back, P. G. O. Freund and M. Forger, "New gravitational instantons and universal spin structures". Phys. Letts. **77B**, 181 (1978).

[8.1] R. P. Geroch, "Spinor structure of space-times in general relativity I and II". J. Math. Phys. **9**, 1739 (1968) and **11**, 343 (1970).
[8.2] J. Milnor, "Spin structures on Manifolds". L'Ens. Math. **9**, 198 (1963).
[8.3] C. J. Isham, "Spinor fields in four-dimensional space-time". Proc. Roy. Soc. **A 364**, 591 (1978).
[8.4] S. J. Avis and C. J. Isham, "Lorentz gauge invariant vacuum functional for quantized spinor fields in non-simply connected space-times". Nucl. Phys. **B 156**, 441 (1979).
[8.5] C. N. Pope and S. W. Hawking, "Generalized spin structures in quantum gravity". Phys. Letts. **73B**, 42 (1978).

[8.6] C. J. Avis and C. J. Isham, "Generalized spin structures on four-dimensional space-times". Commun. Math. Phys. **72**, 103 (1980).

[8.7] T. Kimura, "Divergence of Axial-Vector Current in the Gravitational Field". Prog. Theo. Phys. **42**, 1191 (1969).

[8.8] R. Delbourgo and A. Salam, "The Gravitational correction to PCAC". Phys. Letts. **40B**, 381–2 (1972).

[8.9] T. Eguchi and P. G. O. Freund, "Quantum Gravity and World Topology". Phys. Rev. Letts. **37**, 1251–1254 (1976).

[8.10] P. A. Liggat, and A. J. MacFarlane, "Point Splitting in a Curved Spacetime Background I: Gravitational Contribution to the Axial Anomaly". G.R.G. **10**, 321–324 (1980).

[8.11] N. K. Nielsen, H. Romber and B. Schroer, "Classical Anomalies and Local Version of the Atiyah-Singer Theorem". Phys. Letts. **70B**, 445–448 (1977).

[8.12] N. K. Nielsen, H. Romer and B. Schroer, "Anomalous Currents in Curved Space". Nucl. Phys. **B 136**, 475–492 (1978).

[8.13] J. S. Dowker, "Another discussion of the axial vector anomaly and the index theorem". J. Phys. **A 11**, 347–360 (1978).

[8.14] H. Romer and B. Schroer, "Fractional Topological Configurations and Surface Effects". Phys. Letts. **71B**, 182–184 (1977).

[8.15] C. N. Pope, "Axial Vector Anomalies and the Index Theorem in Charged Schwarzschild and Taub-NUT space". Nucl. Phys. **B 141**, 432–444 (1978).

[8.16] G. W. Gibbons, H. Romer and C. N. Pope, "Index theorem boundary terms for Gravitational Instantons". Nucl. Phys. **B 157**, 337–386 (1979).

[8.17] H. Romer, "Axial anomaly and boundary terms for general spinor fields". Phys. Letts. **83B**, 172–176 (1979).

[8.18] K. Fujikawa, "Path Integral Measure of Gauge Invariant Fermion Theories". Phys. Rev. Letts. **42**, 1195 (1979).

[8.19] D. R. Stump, Phys. Rev. **D 17**, 3281 (1978).

[9.1] W. Pauli, "Lectures on Physics Vol. 1: Selected Topics in Field Quantization" ed. C. P. Enz. M.I.T. Press (1977) p. 33.

[9.2] D. Z. Freedman, P. van Nieuwenhuizen and S. Ferrara. "Progress Towards a Theory of Supergravity". Phys. Rev. **D 13**, 3214 (1976).

[9.3] S. Deser and B. Zumino, "Consistent Supergravity". Phys. Letts. **62B**, 335 (1976).

[9.4] D. Z. Freedman and A. Das, "Gauge Internal Symmetry in Extended Supergravity". Nucl. Phys. **B 120**, 221 (1977).

[9.5] P. K. Townsend, "Cosmological Constant in Supergravity". Phys. Rev. **D 15**, 2808 (1977).

[9.6] S. Deser and B. Zumino, "Broken Supersymmetry and Supergravity". Phys. Rev. Letts. **38**, 1433 (1977).

[9.7] G. W. Gibbons and S. W. Hawking, "Cosmological Event Horizons, Thermodynamics and Particle Creation". Phys. Rev. **D 15**, 2738–2751 (1977).

[9.8] S. J. Avis, C. J. Isham, D. Storey, Phys. Rev. **D 18**, 3568 (1975).

[9.9] T. Yoneya, "Background metric in supergravity theories". Phys. Rev. **D 17**, 2567 to 2572 (1978).

[9.10] S. M. Christensen, M. J. Duff, G. W. Gibbons and M. Rocek, "One loop effects in Supergravity with a Cosmological Content".

[9.11] S. M. Christensen, M. J. Duff, G.W. Gibbons and M. Rocek, "Vanishing One-Loop β Function in Gauged N = 4 Supergravity". Phys. Rev. Letts. **45**, 161–164 (1980).

[9.12] W. Siegel, private communication.

[9.13] E. Cremmer, B. Julia, J. Scherk, "Supergravity in 11 dimensions". Phys. Letts. **76B**, 40 (1978).

[9.14] E. Cremmer and B. Julia, "$N = 8$ Supergravity Theory". Phys. Letts. **80B**, 48 (1978).

[9.15] E. Cremmer and B. Julia, "The SO(8) supergravity". Nucl. Phys. **B 159**, 141–212 (1979).

[9.16] M. J. Duff.

[9.17] S. W. Hawking and C. N. Pope, "Symmetry breaking by Instantons in Supergravity". Nucl. Phys. **B 146**, 381—392 (1978).
[9.18] S. M. Christensen, S. Deser, M. J. Duff and M. T. Grisaru, "Chirality, Self-duality and supergravity counter terms". Phys. Letts. **84B**, 411—415 (1979).
[9.19] R. Kallosh, J.E.T.P. Letts. 192 (1979).
[9.19a] R. Kallosh, "Self-duality in Super Space". Nucl. Phys. **B 165**, 119—128 (1980).
[9.20] S. M. Christensen and M. J. Duff, "New Gravitational Index theorems and super theorems". Nucl. Phys. **B 154**, 301—347 (1979).
[9.21] C. N. Pope.
[9.22] C. J. Isham and A. Chockalingam, "Twisted Super multiplets". I.C.T.P. preprint.

Address: Prof. Dr. G. W. Gibbons
D.A.M.T.P.
University of Cambridge
Silver Street
Cambridge CB 3 9 EW
England

Supergravity and Geometry

P. van Nieuwenhuizen (Stony Brook)

Einstein tried to show that elementary particles are simply singularities of the gravitational field, and a number of people have tried to work that idea out. However, a major problem in this idea is that the singularities of the gravitational field can at best describe bosons, but not fermions. In order to be able to describe also fermions as singularities of some kind of the gravitational field, this field must be fermionic. Supergravity is such a gravitational theory: there are two basic gravitational fields, the first a bosonic gravitational field (the usual vierbein field e_μ^m) and the second a fermionic gravitational field, the gravitino field ψ_μ^a. This gravitino field is a vectorial spinor and carries spin 3/2; it is massless and real. Thus it looks very much like a fermionic counterpart to the vierbein field, and indeed, there is even a symmetry between both: local supersymmetry. Since in supergravity fermions play an essential role, the notion of spin has been unified with the geometry of spacetime. A practical consequence is that relativists who start working in supergravity must use spinors in field theory.

Under local supersymmetry the vierbein field rotates into the gravitino field, and the gravitino field rotates into the covariant derivative of the local gauge parameter (which indicates that the gravitino is in some sense the gauge field of local supersymmetry)

$$\delta e_\mu^m = \frac{\varkappa}{2}\,\bar{\varepsilon}\gamma^m\psi_\mu, \qquad \delta\psi_\mu = \frac{1}{\varkappa}\,\mathrm{D}_\mu\varepsilon \tag{1}$$

These rules are a fusion of two different symmetries of the linearized theory, namely

(i) global (= constant) supersymmetry transformations:

$$\delta e_\mu^m = \frac{\varkappa}{2}\,\bar{\varepsilon}\gamma^m\psi_\mu,$$

$$\delta\psi_\mu = \frac{1}{2}\left(\omega_\mu{}^{mn}(e)\right)_{\text{lin.}}(\sigma_{mn}\varepsilon),$$

see [6].

(ii) local gravitino gauge transformations

$$\delta\psi_\mu = \partial_\mu\varepsilon(x),$$

see [4].

A similar situation holds in Yang-Mills theory:

$$\delta w^a_\mu = D_\mu \Lambda^a$$

is a fusion of

$$\delta w^a_\mu = \partial_\mu \Lambda^a{}_{(x)}$$

and

$$\delta w^a_\mu = \varepsilon^{abc} w^b_\mu \Lambda^c,$$

which are both separate invariances of the linearized action.

Clearly the gauge parameter ε is a fermionic field with spin 1/2; like the gravitino it is real, and in order that the varied fields have the same statistics as the fields themselves, the parameters ε^a ($a = 1,4$ in four-component notation) are anticommuting. The covariant derivative involves the spin connection

$$D_\mu \varepsilon = \partial_\mu \varepsilon + \frac{1}{2}\, \omega_\mu{}^{mn} \sigma_{mn} \varepsilon, \qquad \sigma_{mn} = \frac{1}{4}\, [\gamma_m, \gamma_n] \tag{2}$$

and, as we shall see, the spin connection is a function of the vierbein and gravitino fields. Thus supergravity contains nonvanishing torsion, the torsion tensor being

$$S_{\mu\nu}^{\ \ \varrho} = \frac{\varkappa^2}{4}\, (\overline{\psi}_\mu \gamma^\varrho \psi_\nu) \tag{3}$$

The action of the simplest model of supergravity [1, 2] (appropriately called simple supergravity) is Hilbert's action plus the Rarita-Schwinger action for a real massless spin 3/2 field ψ_μ, the latter appropriately covariantized with respect to gravity

$$\mathscr{L} = -\frac{e}{2\varkappa^2}\, R(e, \omega) - \frac{e}{2}\, \overline{\psi}_\mu \gamma^{\mu\varrho\sigma} D_\varrho \psi_\sigma \tag{4}$$

Here $e = \det e^m_\mu$, $\gamma^{\mu\varrho\sigma}$ is a product of three Dirac matrices antisymmetrized in $\mu\varrho\sigma$ (note that $\gamma^\mu = \gamma^m e^\mu_m$ depends on the vierbein field), while the gravitino curl contains only a Lorentz connection

$$D_\varrho \psi_\sigma = \partial_\varrho \psi_\sigma + \frac{1}{2}\, \omega_\varrho{}^{mn} \sigma_{mn} \psi_\sigma \tag{5}$$

One can now use Palatini formalism, and solve the field equation of the spin connection. The result is

$$\omega_\mu{}^{mn} = \omega_\mu{}^{mn}(e) + \frac{\varkappa^2}{4}\, (\overline{\psi}_\mu \gamma^m \psi^n - \overline{\psi}_\mu \gamma^n \psi^m + \overline{\psi}^m \gamma_\mu \psi^n) \tag{6}$$

where $\omega^{mn}_\mu(e)$ is the usual spin connection without torsion

$$\partial_\mu e^m{}_\nu - \begin{Bmatrix} \varrho \\ \mu\nu \end{Bmatrix} e^m_\varrho + \omega_\mu{}^{mn}(e)\, e_{n\nu} = 0 \tag{7}$$

The torsion (3) follows from (6).

Thus here we have a theory of gravity with a new local symmetry, supersymmetry. The action in (4) is invariant under (1) provided (6) holds. (The proof [and many other details] can be found in a forthcoming Physics Report by the author.) However, this is not yet a theory because the two fields e_μ^m and ψ_μ^a do not form a representation of the symmetries of the theory. These symmetries are

$$\begin{aligned}&\text{(i) general coordinate invariance (reparametrizations)}\\&\text{(ii) local Lorentz invariance}\\&\text{(iii) local supersymmetry,}\end{aligned} \tag{8}$$

and what does not hold is that the commutator of two local symmetries is again a local symmetry. It almost holds, except in one case, namely in the commutator of two local supersymmetry variations of the gravitino [3]. For the vierbein this commutator is given by (use (1))

$$[\delta_{\text{sup}}(\varepsilon_1), \delta_{\text{sup}}(\varepsilon_2)]\, e_\mu^m = \frac{1}{2}\, \bar{\varepsilon}_2 \gamma^m \mathrm{D}_\mu \varepsilon_1 - (1 \leftrightarrow 2) \tag{9}$$

Elementary algebra shows that this commutator is a sum of a general coordinate transformation with parameter $\frac{1}{2}\, \bar{\varepsilon}_2 \gamma^\lambda \varepsilon_1$, a local Lorentz transformation with parameter $-\frac{1}{2}\, \bar{\varepsilon}_2 \gamma^\lambda \varepsilon_1 \omega_\lambda^{mn}(e, \psi)$ and a local supersymmetry transformation with parameter $-\frac{1}{2}\, \bar{\varepsilon}_2 \gamma^\lambda \varepsilon_1 \psi_\lambda^a$

$$\begin{aligned}&\frac{1}{2}\, \bar{\varepsilon}_2 \gamma^m \mathrm{D}_\mu \varepsilon_1 - (1 \leftrightarrow 2) = \xi^\lambda \partial_\lambda e_\mu^m + (\partial_\mu \xi^\lambda)\, e_\lambda^m - \xi^\lambda \omega_\lambda{}^{mn}(e, \psi)\, e_{n\mu} - \xi^\lambda \bar{\psi}_\lambda \gamma^m \psi_\mu,\\&\xi^\lambda = \frac{1}{2}\, \bar{\varepsilon}_2 \gamma^\lambda \varepsilon_1\end{aligned} \tag{10}$$

However, for the gravitino one finds, in addition to three similar terms, extra terms proportional to the field equation of the gravitino. These terms are themselves of course invariances of the action, but they are not one of the three local invariances. If one would add this invariance to the list in (8) and keep computing commutators, one would find an infinite dimensional algebra.

One can find a representation of the local algebra in (8) in terms of the following fields [4]

$$(e_\mu^m, \psi_\mu{}^a, S, P, A_m) \tag{11}$$

In the action one simply must add $-\frac{\varkappa^2}{3}(S^2 + P^2 - A_m^2)$ so that the scalar S, and pseudoscalar P, and the axial vector A_m are auxiliary fields. The transformation rules of the fields in (11) are not very complicated, but will not be given here.

We mentioned already the Palatini formalism, and one might wonder whether one can treat ω_μ^{mn} as an independent field. Indeed, one can do so, but the transformation law of $\omega_\mu^{mn}(e, \psi)$ in second order formalism [1] is not the same as $\delta\omega_\mu^{mn}$ in first order formalism [2] (even after substituting (6) in the result for $\delta\omega_\mu^{mn}$).

Actually — one could have anticipated this. The difference of the gravitino law in first and second order formalism is given by (see [1])

$$\delta\psi_\mu(\text{first}) - \delta\psi_\mu(\text{second}) = \varkappa^{-1}[\omega_\mu^{mn} - \omega_\mu^{mn}(e, \psi)] \left(\frac{1}{2}\,\sigma_{mn}\varepsilon\right) \tag{12}$$

and is thus proportional to the field equation of the spin connection, see (6). In the same way one finds that

$$\delta\omega_{\mu mn}(\text{second}) - \delta\omega_{\mu mn}(\text{first}) = \frac{\varkappa}{4}\,(\bar{\varepsilon}\gamma_n R_{\mu m} - \bar{\varepsilon}\gamma_m R_{\mu n} - \bar{\varepsilon}\gamma_\mu R_{mn}) \tag{13}$$

where $R_{\mu m}$ is proportional to the gravitino field equation

$$\begin{aligned} R_{\mu m} &= e_m{}^\nu(\mathrm{D}_\mu\psi_\nu - \mathrm{D}_\nu\psi_\mu + \varepsilon_{\mu\nu}{}^{\varrho\sigma}\mathrm{D}_\varrho\psi_\sigma) = -\gamma_\lambda\sigma_{\mu m}R^\lambda \\ R^\lambda &= \gamma^{\lambda\varrho\sigma}\mathrm{D}_\varrho\psi_\sigma = \text{gravitino field equation} \end{aligned} \tag{14}$$

So, in going from second to first order formalism one must add

$$\delta\psi_\mu{}^{(\text{extra})} \sim \frac{\delta I}{\delta\omega}, \qquad \delta\omega_\mu{}^{mn(\text{extra})} \sim \frac{\delta I}{\delta\psi} \tag{15}$$

and it is clear that in the variation of the action such terms can cancel

$$\delta I = \frac{\delta I}{\delta e} + \frac{\delta I}{\delta\psi}\,\delta\psi + \frac{\delta I}{\delta\omega}\,\delta\omega \tag{16}$$

In all formulations of supergravity, it is the second order formalism which seems the appropriate formulation. For example, no auxiliary fields are known for the first order formalism (the analogue of (11) would need, in addition to ω_μ^{mn}, quite a number of auxiliary fields), while in superspace second order formalism is the only possible formalism.

The observant reader may have become uneasy at this point. The parameters in the commutator in (10) are field-dependent. Phrased differently, instead of structure constants, we have structure functions. However, one can find a larger set of auxiliary fields (larger than in (11)) for which one has constant structure constants. This set follows from superspace, to which we now turn. However, there is nothing bad about structure functions; it is merely unusual.

The variation of the action in (4) is invariant under the local symmetries in (8), but the Lagrangian varies as follows

$$\delta\mathscr{L} = \partial_\lambda(\xi^\lambda\mathscr{L}) + \partial_\lambda\left(-\frac{1}{8}\,\bar{\varepsilon}\gamma^\lambda\gamma_\mu R^\mu\right) \tag{17}$$

The first term is of course the standard result for the variation of a scalar density, and shows that general coordinate transformations are not internal symmetries, but rather spacetime symmetries, in contrast to local Lorentz transformations which are genuine internal symmetries and which keep even the Lagrangian invariant. The last term reveals a deep property: local supersymmetry is more like general coordinate than like local Lorentz transformations, and is thus in some sense a spacetime symmetry. Going back to (10), we see that two local supersymmetry transformations produce a general coordinate transformation (plus more) so that in this sense supergravity is the square root of general relativity. More precisely: in Dirac's constrained Hamiltonian formalism the generators $\mathscr{H}^m$, Q^α and J^{mn} for the symmetries in (8) satisfy Poisson bracket relations, one of which reads [5]

$$\{Q^\alpha, Q^\beta\} = \frac{1}{2} (\gamma_m C^{-1})^{\alpha\beta} \mathscr{H}^m$$

where $\mathscr{H}^m$ is a covariant translation and generates the transformations in (10) in Hamiltonian formalism.

If local supersymmetry is a spacetime symmetry, what is then that spacetime? Just as one associates with ξ^λ a coordinate x^λ it is tempting to associate with ε^a a coordinate θ^a. Thus one is led to superspace: a space with four bosonic and four anticommuting fermionic coordinates [6]. From a more mathematical point of view, one can consider the global algebra which underlies local supersymmetry. It is the super Poincaré algebra, with generators Q^α, P_m, J_{mn}. In addition to the usual Poincaré algebra one has two relations which say that Q^α is a conserved spin 1/2 generator

$$[Q^\alpha, P_m] = 0, \qquad [Q^\alpha, M_{mn}] = \frac{1}{2} (\sigma_{mn})^\alpha{}_\beta Q^\beta \tag{18}$$

and the equivalent of (18)

$$\{Q^\alpha, Q^\beta\} = \frac{1}{2} (\gamma^m C^{-1})^{\alpha\beta} P_m, \qquad C\gamma^m C^{-1} = -(\gamma_m)^T \tag{19}$$

That this is a closed algebraic system follows from the fact that the Jacobi identities are satisfied,

$$\big[A, [B, C]\big] = \big[[A, B], C\big] + (-)^{ab} \big[B, [A, C]\big], \tag{20}$$

where $[A, B]$ is a commutator unless both A and B are Q generators, in which case one has an anticommutator ($a = 0$ for bosonic generators and $a = 1$ for fermionic generators).

One can now consider coset spaces $(P + M + Q)/M$, and identify x^μ and θ^α with the coset generators P_m and Q^α. In this way one can find, using standard mathematical techniques, general properties of superspace. In particular, it is clear that local Lorentz transformations are the internal symmetries in superspace, but one also finds that *fermionic general coordinate transformations are equivalent to local supersymmetry.*

An elaborate theory of superspace supergravity has been developed. In the most obvious extension of general relativity [7], one introduces "achtbeins"

$$E_\Lambda{}^M, \qquad \Lambda = (\mu = 1, 4; \alpha = 1, 4), \qquad M = (m = 1, 4;\ a = 1, 4) \tag{21}$$

and spin connections whose base manifold index is again 8-dimensional and which gauge the internal symmetry (only local Lorentz invariance)

$$h_A{}^{mn}, \qquad A = (\mu = 1, 4;\ \alpha = 1, 4); \qquad m, n = 1, 4 \tag{22}$$

Covariant derivatives are defined by $D_A = \partial_A + \frac{1}{2} h_A{}^{mn} X_{mn}$ where X_{mn} are the Lorentz generators, and defining flat covariant derivatives by $D_A \equiv E_A{}^A D_A$, one defines supertorsions and supercurvatures by

$$[D_A, D_B] = -2T_{AB}{}^C D_C + R_{AB}{}^{mn} \frac{1}{2} X_{mn} \tag{23}$$

The symmetries in superspace are

(i) general supercoordinate invariance

$\left(\xi^A(x, \theta) = \xi^\mu, \xi^\alpha\right)$

(ii) local Lorentz invariance $\left(\Lambda^{mn}(x, \theta)\right)$

To make contact with ordinary space one identifies $\xi^\mu(x, \theta = 0)$ with the ordinary reparametrization parameter, $\xi^\alpha(x, \theta = 0)$ with the supersymmetry parameter $\varepsilon^\alpha(x)$ and $\Lambda^{mn}(x, \theta = 0)$ with the local Lorentz parameter $\lambda^{mn}(x)$. Further, $h_\mu^{mn}(x, \theta = 0) = \omega_\mu^{mn}$, $E_\mu^m(x, \theta = 0) = e_\mu^m$ and $E_\mu^a(x, \theta = 0) = \psi_\mu^a$. In superspace the transformation rules are superreparametrization and local Lorentz invariance

$$\delta E_A{}^M = \xi^\pi \partial_\pi E_A{}^M + (\partial_A \xi^\pi)\, E_\pi{}^M + \frac{1}{2} \Lambda^{mn} X_{mn} E_A{}^M \tag{24}$$

If one requires that these rules for $E_\mu^m(x, \theta = 0)$ and $E_\mu^a(x, \theta = 0)$ are compatible with the transformations of the ordinary space approach, one finds that the different coefficients of powers of θ in $E_A{}^M$, $h_A{}^{mn}$, ξ^A and Λ^{mn} are functions of the fields and parameters of the x-space approach. In particular, one can

(i) Determine $E_A{}^M$ to order θ and $h_A{}^{mn}$ to order $\theta = 0$. This is due to the transport term $\xi^\alpha \partial_\alpha E_A{}^M$ in (24).

(ii) Substituting these results into (23), one finds that certain components of $T_{AB}{}^C$ at order $\theta = 0$ vanish.

(iii) Since $T_{AB}{}^C$ are tensors in superspace, this implies that these supertorsion components vanish to all order in θ. (Just like a tensor in general relativity which vanishes at the origin of any coordinate system, is identically zero.)

In this way one finds the following constraints on the supertorsions of simple supergravity [7] (a, b flat fermionic indices; m, n flat bosonic indices).

$$T_{ab}{}^c = T_{am}{}^n = T_{mn}{}^r = 0, \qquad T_{ab}{}^m = (C\gamma^m)_{ab} \tag{25}$$

One can solve these constraints [8] and finds then three sets of results

(i) the spin connection $h_A{}^{mn} = (h_\mu{}^{mn}, h_\alpha{}^{mn})$ becomes a function of the supervielbein $E_A{}^M$. This is the equivalent of second order formalism in the x-space approach.

(ii) the bosonic components $E_m{}^A$ of the inverse supervielbein are functions of the fermionic components $E_\alpha{}^A$. This is thus "more second order formalism".

(iii) the fermionic components $E_\alpha{}^A$ are functions of prepotentials, just like in electromagnetism the constraint $\varepsilon^{\mu\nu\varrho\sigma}\partial_\nu F_{\varrho\sigma} = 0$ can be solved to yield the potential A_σ in $F_{\varrho\sigma} = \partial_\varrho A_\sigma - \partial_\sigma A_\varrho$. In fact, just as in this example, one finds also that supergravity, when formulated in terms of prepotentials, has an extra local gauge invariance under which the $E_A{}^M$ are inert.

The constraints for simple supergravity in (25) were found by building a bridge between ordinary space and superspace (gauge completion). For the extended supergravities (theories with more than one gravitino) it is very hard to find the correct set of constraints. In fact, it was found that in $N = 2$ extended supergravity[1]) the constraints involve squares of $T_{AB}{}^C$ [10]. One must impose these constraints in order to be able to write down an action. In fact, the action is just the invariant volume element [7] in the $N = 1$ theory

$$I = \int \mathrm{d}^4x\, \mathrm{d}^4\vartheta \text{ (superdeterminant } E_A{}^M) \tag{26}$$

Since one varies under the constraints (25), one does not find $E_A{}^M = 0$, but rather the same field equations as obtained from (4). The constraints are part of the "kinematics" of the theory, and hold, whether or not one chooses the dynamics as in (26) or as given by some other action.

It is unsatisfactory that the constraints are imposed from the outside. It would be nicer to have them follow as field equations from some action. This is what another geometrical approach aims at: the group manifold approach [11]. In this approach one has a larger space than superspace, namely 14 instead of 8 dimensional: one coordinate per generator in (18, 19). The field equations then tell one that fields do not depend on the Lorentz coordinates, and hopefully this method will yield a derivation of the constraints in (25) which allows one to find the auxiliary fields of all supergravity theories. In the group manifold itself one has a closed gauge algebra (in fact, one has so many auxiliary fields that one could almost speak of the maximal set of auxiliary fields), but in using all field equations one looses closure. The standard problem is thus to understand which field equations to use and where to stop, and this is the cardinal problem in this approach.

The theories in supergravity cannot have more than 8 gravitinos, because the irreducible representations of the algebra in (18, 19) with N $Q^{\alpha,i}$ ($i = 1, N$) involve states with spins exceeding 2 as soon as $N > 8$. It has been shown by Berends, van Holten, de Wit and the author [12] that one cannot couple spin 5/2 in a consistent way to gravity. Several authors have tried to improve this result, without any success. Thus nature seems to allow as fundamental particles only those whose spin J is not larger than 2. For supergravity this means that only the largest model (the $N = 8$ model) can come near being "the model of the world" (all other models with $N < 8$ have too few particles to be realistic. See, however, Ferrara's talk).

---

[1]) This theory unifies electromagnetism and gravity: the photon, graviton and two gravitinos rotate into each other under local supersymmetry [9].

Supergravity has remarkable quantum properties. Whereas the coupling of Einstein gravity to *any* matter always leads to a divergent S-matrix (and the divergences are nonrenormalizable because they are of a different functional form than the original action), in supergravity all infinities in the S-matrix cancel "miraculously." In fact, this is one way in which supergravity might have been discovered. The first breakthrough in quantum gravity occurred in the $N = 2$ model [9] where a calculation of photon-photon scattering turned out to be finite [13] at the one-loop level. At the two-loop, level, it was shown by theoretical considerations [14] that supergravity is finite. At the three-loop level an invariant exists which, if it occurs in the S-matrix, would destroy finiteness of the theory [15]. However, I believe that this 3-loop invariant is a red herring; in fact, quite recently Grisaru, Rocek and Siegel found by superspace calculations that a model in global supersymmetry which looks very much like $N = 8$ supergravity is 3-loop finite, although also here a possible 3-loop invariant exists [16]. Clearly — if supergravity is all loop finite, it would be an interesting model for quantum gravity.

Let me now mention a few separate results.

Inequivalence of different field representations for given spin. One can represent a spin 0 field either by Φ or by $A_{\mu\nu} = -A_{\nu\mu}$; in the latter case the action reads

$$L = -(\partial_\mu A_{\nu\varrho} + \text{cyclic terms})^2 \tag{27}$$

Coupling to gravity, and adding as gauge fixing term for the local invariance $\delta A_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ the deltafunction $\delta(\mathrm{D}^\mu A_{\mu\nu} - b_\nu)$, one finds that b_ν is constrained by $\mathrm{D}^\nu b_\nu = 0$ because $\mathrm{D}^\nu \mathrm{D}^\mu A_{\mu\nu} \sim R^{\mu\nu} A_{\mu\nu} = 0$. Keeping track of all necessary ghosts [17], one finally finds that

(i) the S-matrix for Φ coupled to gravity equals the S-matrix for $A_{\mu\nu}$ coupled to gravity [18].

(ii) the trace anomalies differ [19],

$$\Delta L(A_{\mu\nu} + \text{gravity}) - \Delta L(\Phi + \text{gravity}) = -\chi \tag{28}$$

where χ is the Euler topological invariant. This result will be further discussed by Duff.

R^2 theories without ghosts with propagating torsion [20]. Taking the spin connection as an independent field, one can consider actions with only two derivatives but with propagating spin connection. These theories can be made unitary by fixing the free parameters in such a way that the residues of all propagators are positive. However, none of these theories is renormalizable.

All existing R^2 theories were diagnosticized and found to violate unitarity. Thus it seems that R^2 theories with propagating torsion are sick, at least if one considers ordinary perturbation theory. They are either nonrenormalizable (not: finite) or nonunitary.

References

[1] D. Z. Freedman, P. van Nieuwenhuizen, S. Ferrara, Phys. Rev. **D 13**, 3214 (1976).
[2] S. Deser and B. Zumino, Phys. Lett. **62B**, 335 (1976).
[3] D. Z. Freedman and P. van Nieuwenhuizen, Phys. Rev. **D 14**, 912 (1976).
[4] S. Ferrara and P. van Nieuwenhuizen, Phys. Lett. **74B**, 333 (1978).
K. Stelle and P. C. West, Phys. Lett. **74B**, 330 (1978).
[5] C. Teitelboim, Phys. Rev. Lett. **38**, 1106 (1977).
M. Pilati, Princeton Ph. D., Jan. 1980.
[6] A. Salam and J. Strathdee, Nucl. Phys. **B 76**, 477 (1974).
[7] J. Wess and B. Zumino, series of papers in Physics Letters.
[8] S. J. Gates and W. Siegel, Nucl. Phys. **B 163**, 519 (1980).
[9] S. Ferrara and P. van Nieuwenhuizen, Phys. Rev. Lett. **37**, 1669 (1976).
[10] L. Castellani, J. Gates and P. van Nieuwenhuizen, Phys. Rev. **D 22**, 2364 (1980).
[11] A. D'Adda, R. D'Auria, P. Fré, T. Regge, Rivista del Nuovo Cimento **6** (1980).
R. D'Auria, P. Fré, T. Regge, Rivista del Nuovo Cimento, in press.
[12] F. A. Berends, J. W. van Holten, B. de Wit and P. van Nieuwenhuizen, J. Phys. **A 13**, 1643 (1980), Phys. Lett. **83B**, 188 (1979) and Nucl. Phys. **B 154**, 261 (1979).
[13] M. T. Grisaru, P. van Nieuwenhuizen and J. A. M. Vermaseren, Phys. Rev. Lett. **37**, 1662 (1976).
[14] M. T. Grisaru, Phys. Lett. **66B**, 75 (1977).
[15] S. Deser, J. Kay, K. Stelle, Phys. Rev. Lett. **38**, 527 (1977).
[16] M. T. Grisaru, M. Rocek, W. Siegel, Phys. Rev. Lett. **45**, 1063 (1980).
[17] W. Siegel, Phys. Lett. **93B**, 170 (1980).
[18] E. Sezgin and P. van Nieuwenhuizen, Phys. Rev. **D 22**, 301 (1980).
[19] M. J. Duff and P. van Nieuwenhuizen, Phys. Lett. **94B**, 179 (1980).
[20] D. Neville, Phys. Rev. **D 21**, 867 and 2075 (1980).
E. Sezgin and P. van Nieuwenhuizen, Phys. Rev. **D 21**, 3269 (1980).

Address: Prof. Dr. P. van Nieuwenhuizen
Institute of Theoretical Physics
State University of New York at Stony Brook
Long Island, N.Y. 11794
USA

Aspects of Supergravity Theories

S. Ferrara (Geneva)

Abstract

Some features of supergravity as a unitary gauge field theory for all fundamental particle interactions are reviewed. Special emphasis is placed on the multiplet structure of massive and massless representations of extended supersymmetry, and their symmetry properties are discussed.

1. A Framework for a Unitary Field Theory

The present theoretical understanding of low energy phenomena seems to indicate that gauge field theory is a viable framework for the description of all fundamental forces of Nature. This is true for long-range electromagnetic and gravitational forces as well as for weak and strong short-range nuclear forces. If one accepts such a dynamical framework for the known interactions of the fundamental constituents of matter, there are several motivations which suggest the use of a further economical principle, that is, different low energy symmetries become unified at energies which are much higher than the energy scales needed today to understand low energy physics. In the framework of gauge quantum field theory, a possible scenario for unification of the non-gravitational interactions is provided by GUTs [1] (Grand Unified Theories). This picture embeds in a single gauge theory the Glashow-Weinberg-Salam $SU(2)_{left} \otimes U(1)_{el.}$ theory of electroweak interactions and the colour SU(3) gauge theory (quantum chromodynamics) of strong interactions. The minimal GUT is given by the Georgi-Glashow SU(5) model [2] which provides the minimal embedding of $SU(2)_{left} \otimes U(1)_{el.} \otimes SU(3)_{colour}$ in a simple group. This model, as well as any other GUT [3], must recover the fact that electromagnetic and weak forces are unified at 100 GeV. This implies, using renormalization group arguments, that the grand unification scale is of the order of 10^{15} GeV. GUTs have by now had much success, for example, the explanation of charge quantization, the prediction of the low energy weak angle θ_W and some relations between quarks and leptons. They also imply new physical phenomena which can be tested experimentally like the proton decay rate and neutrino masses [4]. However, GUTs have serious drawbacks because they have an intrinsic hierarchy problem, they do not explain the family repetition and, last but not least, the fact that, although they become true symmetries at scales not far from the Planck mass 10^{19} GeV, they completely neglect gravity.

On the other hand, any attempt at superunification must incorporate a gauge field, the metric tensor, which describes the gravitational interaction. This field, in the quantum theory, describes a new gauge quantum, the graviton, which exists in two states of helicity $\lambda = \pm 2$. This gauge field is related to the local symmetry of the Einstein Lagrangian, i.e., the gauge Poincaré symmetry.

The main motivation of supergravity [5, 6] is to provide a possible scenario for a superunified gauge theory having the Planck scale as a unification scale. It is, in fact, clear that in any attempt to construct a superunified gauge quantum field theory, the unifying gauge principle must contain the space-time symmetry and the internal symmetries in a unique algebraic structure. More importantly, in a truly unified gauge theory, all the interactions should become purely geometrical and any distinction between matter and gauge fields should be elusive. Because of the diffferent spin and statistics properties of the basic constituents of matter (fermions and bosons) this seems to require that the unifying algebraic structure must have the property of converting bosons into fermions and vice versa. This is achieved by Graded Lie Algebras (GLAs) or supersymmetries [7, 8], which have a multiplication rule which contains both commutators and anticommutators.

Supergravity is a synonym for the dynamical theory of local (gauge) supersymmetry. In supersymmetry, the odd part (fermionic generators) of the GLA contains N spinorial generators of Majorana type Q_α^i ($\alpha = 1, \ldots, 4$, $i = 1, \ldots, N$). These generators form the so-called grading representation of the Lie algebra part of GLA which contains the Pioncaré algebra with the possible addition of an internal symmetry acting on the index i. The representation properties of the spinorial generators are[1])

$$[Q_\alpha^i, M_{\mu\nu}] = i(\sigma_{\mu\nu})_\alpha^\beta\, Q_\beta^i, \qquad [Q_\alpha^i, P_\mu] = 0 \tag{1}$$

and their anticommutation relations are

$$\{Q_\alpha^i, \bar{Q}_\beta^j\} = -2\gamma_{\alpha\beta}^\mu P_\mu \delta^{ij} + Z^{ij}\delta_{\alpha\beta} + \tilde{Z}^{ij}\gamma_{\alpha\beta}^5 \tag{2}$$

The generators Z^{ij}, $\tilde{Z}^{ij}$, both antisymmetric in the i, j indices, belong to the centre of the GLA and for this reason are called central charges [9]. Although these generators become important in some models, it is consistent to set them equal to zero in (1) and (2) without violating any of the Jacobi identities. From now on we will mostly consider the algebra defined by (1) and (2) with $Z^{ij} = \tilde{Z}^{ij} = 0$.

If we invert Eq. (2) we get

$$P_\mu = -\frac{1}{4N}\,\bar{Q}^i\gamma_\mu Q^i \tag{3}$$

which shows that the spinorial generators are, in a certain sense, the square roots of translations. From the previous relation it is evident that, if we gauge the fermionic generators Q_α^i, i.e., if we make a theory invariant under supersymmetry transformations with space-time dependent (anticommuting) parameters $\varepsilon_\alpha^i(x)$, then this theory must also be invariant under general co-ordinate transformations. Conversely, any theory

[1]) In the present review we often use different metric conventions according to the original papers quoted in the references.

with general covariance and global supersymmetry is also locally supersymmetric. When local symmetry transformations with parameter $\varepsilon^i_\alpha(x)$ are introduced, then one expects gauge fields to be necessary in order to make a given theory locally supersymmetric. This is entirely analogous to the introduction of the photon in order to extend the global (first kind) phase invariance of the Dirac theory to a local (second kind) invariance. This gauge field must, in the case of supersymmetry, transform as $\partial_\mu \varepsilon^i_\alpha$ (+ more terms) and is therefore represented by a Rarita-Schwinger field $\psi^i_{\mu\alpha}(x)$. This fermionic gauge field is supposed to describe, in the absence of supersymmetry breaking, a new massless particle of helicity states $\lambda = \pm 3/2$, the gravitino. The existence of a new gauge quantum of half-integral spin is the very clue to supergravity. This hypothetical particle of spin 3/2 is, in fact, the bridge between the space-time symmetry whose gauge quantum has spin two (graviton) and internal symmetries whose gauge quanta have spin one (Yang-Mills vector bosons). Moreover, from the multiplet structure of massless representations of extended ($N > 1$) supersymmetry [10] it follows that gauge particles of spin 2, 3/2 and 1 can also be unified with particle fields of spin 1/2 and 0. Then, in supergravity theories, two fundamental goals are simultaneously achieved: unification of space-time and internal local symmetries in a single gauge theory and unification of gauge fields and matter fields in a single irreducible representation of the underlying symmetry group.

2. Supergravity from first principles

Quite independently of the unification programme several motivations have been given for the introduction of supergravity in particle physics.

Firstly, to promote supersymmetry to a local invariance. This requires, at the same time, invariance under general co-ordinate transformations. The gauging of supersymmetry, irrespective of the particular mathematical framework which has been used, yields uniquely the same action [5, 6].

When the gauge invariance of the free massless Rarita-Schwinger field is extended to an interacting theory, supergravity emerges as the only solution for a consistent coupling to other particle fields, overcoming the old difficulties for coupled higher spin fields equations [11].

Supergravity provides, for the first time, the idea that fermionic fields, which are usually associated with matter, can be genuine gauge fields. This gives a subtle interplay between space-time geometry and the quantum mechanical concept of spin.

Supergravity was also motivated as an attempt to build a meaningful quantum theory of gravitation. Indeed, the improvement of the ultra-violet behaviour with respect to Einstein's theory, due to the short-distance contributions of the supersymmetric partners of the graviton, is spectacular [12]. The ultimate hope is that some versions of supergravity, the present preference being the maximally extended $N = 8$ theory [13], will lead to a finite theory of quantum gravity.

Supergravity gives a natural justification as to why gravity should be quantized [14]. In fact, if there are symmetry operations which mix the metric tensor field with other particle fields it would be perverse to quantize all fields but the metric. For instance, in the simplest supergravity theory, the fact that the Rarita-Schwinger field is a quantum field, already suggests that the metric tensor should be quantized as well.

There have been several elegant reformulations of the basic ($N = 1$) supergravity theory [5, 6]. In first order form, $N = 1$ supergravity can be regarded as the Einstein-Cartan theory for a spin 3/2 massless particle with the non-minimal substitution $\partial_\mu \to \partial_\mu + 1/2\, \omega_\mu{}^{ab}\sigma_{ab}$ in the Rarita-Schwinger Lagrangian [6]. This substitution is not minimal in the sense that it is not the covariant derivative which acts on a spin 3/2 field. In the Lagrangian this makes a difference due to contorsion terms, but it is necessary in order to preserve the fermionic gauge invariance. Supergravity has also been derived as the gauge theory of the graded Poincaré group [15] or de Sitter group [16] in a space-time supplemented by a torsion-free condition. More sophisticated techniques have also been used to derive supergravity in a simpler way [17, 18]. One of them uses the concept of superspace which is [17, 19, 21] the quotient space of the graded Poincaré group over the Lorentz group. In superspace, multiplets of fields are described by a single field, the superfield. In spite of these significant technical improvements we would like to go through a derivation of the theory which does not require any knowledge of differential geometry and group theory and is based on some very simple physical considerations. From the representation theory of global supersymmetry it is known that a massless (Majorana) particle with helicity $\lambda = \pm 3/2$ can form a supermultiplet with a bosonic partner of helicity ± 1 or ± 2. Indeed in a free field theory both choices are equally possible. The first choice is the supergravity multiplet (± 2, $\pm 3/2$) with free Lagrangian:

$$\mathscr{L}^0 = \mathscr{L}^0_{\text{Einstein}}(h_{\mu\nu}) - \frac{1}{2}\, \varepsilon^{\mu\nu\varrho\sigma}\overline{\psi}_\mu\gamma_5\gamma_\nu\, \partial_\varrho\psi_\sigma \tag{4}$$

$\mathscr{L}^0_{\text{Einstein}}$ is the linearized Einstein Lagrangian with $g_{\mu\nu} = \eta_{\mu\nu} + kh_{\mu\nu}$. $\mathscr{L}^0$ is invariant under two separate Abelian transformations

$$\begin{aligned} \delta h_{\mu\nu} &= \partial_\mu\xi_\nu + \partial_\nu\xi_\mu \\ \delta\psi_\mu &= \partial_\mu\alpha \end{aligned} \tag{5}$$

and global supersymmetry rotation

$$\begin{aligned} \delta h_{\mu\nu} &= \bar{\varepsilon}\gamma_\mu\psi_\nu + \bar{\varepsilon}\gamma_\nu\psi_\mu \\ \delta\psi_\mu &= \partial_\varrho h_{\mu\sigma}\sigma^{\varrho\sigma}\varepsilon \end{aligned} \tag{6}$$

The alternative choice is the multiplet ($\pm 3/2$, ± 1) with free Lagrangian

$$\mathscr{L}^1 = -\frac{1}{4}\, F^{\mu\nu}F_{\mu\nu} - \frac{1}{2}\, \varepsilon^{\mu\nu\varrho\sigma}\overline{\psi}_\mu\gamma_5\gamma_\nu\partial_\varrho\psi_\sigma \tag{7}$$

$\mathscr{L}^1$ is invariant under two separate Abelian transformations

$$\delta A_\mu = \partial_\mu \Lambda, \qquad \delta\psi_\mu = \partial_\mu \alpha \qquad (F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu) \tag{8}$$

and a global supersymmetry rotation

$$\delta A_\mu = \bar{\varepsilon}\psi_\mu, \qquad \delta\psi_\mu = \sigma_{\varrho\sigma}\gamma_\mu F^{\varrho\sigma}\varepsilon \tag{9}$$

However, the difference between $\mathscr{L}^0$ and $\mathscr{L}^1$, which correspond to the possible embedding of the gravitino in a supermultiplet, comes when we try to promote the two theories described by (4) and (7) to a fully interacting theory. If we have $\varepsilon = \varepsilon(x)$, i.e., we perform a local supersymmetry transformation, then a new interaction term is required of the form $k\bar{\psi}_{\mu\alpha}J^{\mu\alpha}$ (k is the gravitational coupling constant) with $\partial^\mu J_{\mu\alpha}(x) = 0$. This is nothing but the Noether coupling. On the other hand, it turns out that, under a supersymmetry variation, the spinor supersymmetry current transforms into the stress tensor of the system $T^{\mu\nu}$. Hence the $k\bar{\psi}_\mu J^\mu$ coupling requires, at the same time, a $kh_{\mu\nu}T^{\mu\nu}$ term and therefore only the ansatz given by (4) is possible. The final theory derived by this step-by-step procedure gives the supergravity Lagrangian in the form

$$\mathscr{L}_{SG} = -\frac{1}{2k^2}\sqrt{-g}\,R(g_{\mu\nu}) - \frac{1}{2}\,\varepsilon^{\mu\nu\varrho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu \mathrm{D}_\varrho\psi_\sigma$$

$$-\frac{e}{32}\,k^2[(\bar{\psi}^\mu\gamma^\nu\psi^\varrho)\,(\bar{\psi}_\mu\gamma_\nu\psi_\varrho + 2\bar{\psi}_\nu\gamma_\mu\psi_\varrho - 4(\bar{\psi}_\mu\gamma\psi)^2] \tag{10}$$

$g_{\mu\nu} = e_\mu{}^a e_{\nu a}$ and $e_{\mu a}$ is the vierbein field. $\mathscr{L}_{\mathrm{SG}}$ is invariant under the following (non-Abelian) gauge transformation with $\varepsilon = \varepsilon(x)$:

$$\delta e_{a\mu} = k\bar{\varepsilon}\gamma_a\psi_\mu$$

$$\delta\psi_\mu = \frac{2}{k}\,\mathrm{D}_\mu\varepsilon + \frac{1}{4}\,k\sigma^{ab}\varepsilon(2\bar{\psi}_\mu\gamma_a\psi_b + \bar{\psi}_a\gamma_\mu\psi_b) \tag{11}$$

D_μ is the ordinary gravitational derivative with Christoffel affinity. Then it happens that, at the free level ($k = 0$), there are two independent Abelian transformation (5) and a global (non-Abelian) supersymmetry rotation (6). At the coupled level (10) they become a single non-Abelian gauge transformation (11). Thus the law $\partial\psi_\mu = (2/k)\,\mathrm{D}_\mu\varepsilon$ ($\mathrm{D}_\mu = \partial_\mu +$ contorsion terms) is of the same form as the transformation law of a Yang-Mills potential A_μ^a under a local Yang-Mills transformation $\delta A_\mu^a = \mathrm{D}_\mu\Lambda^a = \partial_\mu\Lambda^a + f^{abc}\Lambda^b A_\mu^c$. The only difference is that in supergravity ω_μ^{ab}, the spin connection is a non-linear function of the field variables. Finally, there is a very elegant derivation [22] of the supergravity Lagrangian (10) which does not even require any knowledge of the transformation laws of the fields given by (11), but only the fact that the gravitino field describes two physical massless states of helicities $\pm 3/2$. This derivation is straightforward and only requires the knowledge of the gravitational Born amplitude for the scattering of two spin 3/2 particles. It also emphasizes the interpretation of the four-fermion coupling present in (10) as a seagull term of the same nature as similar terms present in ordinary Yang-Mills theories and in scalar electrodynamics. If we take the Born amplitude for the scattering of two spin 3/2 fields through a one-graviton exchange, the polarization tensor $\psi_\mu(p)$ of any external leg of momentum p_μ on the mass-shell $\big(p^2\psi_\mu(p) = 0\big)$ satisfies the equations

$$p \cdot \psi(p) = 0\,, \qquad \not{p}\psi_\mu(p) = p_\mu \gamma \cdot \psi(p) \tag{12}$$

Equations (12) reduce the number of physical components of $\psi_\mu(p)$ to four. To reduce them further to 2 the S matrix element must vanish when we make the substitution $\psi_\mu(p) = \varepsilon p_\mu$ where ε is a constant Majorana spinor. It is a simple exercise to prove that the Born amplitude due to one-graviton exchange does not fulfil this requirement and that the S matrix vanishes only if the contact term given in (10) is added to the Born amplitude. There is a simple argument which shows that this extra term can only be a four-fermion coupling and that additional contact terms with more spin 3/2 fields would never help. Let us consider the scattering, in the tree approximation, of n gravitini. The scattering can then proceed through exchanges of gravitons between gravitini (tri-linear coupling). The gravitational coupling constant k in such graphs always appears with power $2(n-1)$. Let us now assume that we replace one of the external polarization tensors by its momentum, compute the sum of all graphs contributing to the S matrix and find that it does not vanish. This means that one has to introduce an additional term in the Lagrangian of the form $g_{2n}\partial^m(\bar{\psi}\psi)^n$ where ∂^m means m derivatives and indices have been omitted. Now, a trivial dimensional argument restricts such contact terms very severely: from the kinetic term of the action the dimension of ψ_μ is $(D-1)/2$ in unit of mass, D being the space-time dimension which we leave arbitrary. Then, the dimension of the coupling g_{2n} is:

$$[g_{2n}] = D - n(D-1) - m$$

However, such contact terms, if required, should match with possibly non-gauge invariant terms, and for them $g_{2n} \sim k^{2(n-1)}$. Since $[k] = (2-D)/2$ then we get the consistency equation

$$2(n-1)\,\frac{2-D}{2} = D - n(D-1) - m \quad \text{i.e.} \quad n = 2 - m$$

The only possible solution is $m = 0$, $n = 2$ for any D. We conclude that the contact term for supergravity, in any space-time dimension, can only be a four-fermion coupling with no derivatives.

3. Symmetries of extended supermultiplets

In any supersymmetric theory particle fields are grouped in supermultiplets. These supermultiplets are collections of ordinary fields with different spin, statistics and internal symmetry properties. In this section we will describe, in some detail, the structure of massive and massless multiplets of N extended supersymmetry and their internal symmetry properties. The states of these multiplets are supposed to be described by the asymptotic fields of supersymmetric quantum field theories. According to Salam and Strathdee [23, 24], the construction of particle supermultiplets uses the Wigner method of induced representations. We will temporarily consider the supersymmetry algebra in (2) in the absence of central charges Z^{ij} and $\bar{Z}^{ij}$ (zero charge sector). We first consider the stability subalgebra of a time-like momentum $p^\mu = (M, \vec{0})$. M is the common mass of different spin states in a given massive supermultiplet. In the Majorana representation the stability subalgebra becomes ($M = 1$)

$$\{Q^i_\alpha, Q^j_\beta\} = \delta_{\alpha\beta}\delta^{ij} \qquad (\alpha, \beta = 1, \dots 4; \quad i, j = 1, \dots N) \tag{13}$$

The anticommutators (13) define the Clifford algebra for the group SO($4N$). Its unique irreducible representation has dimension 2^{2N} and it breaks into two inequivalent irreducible spinor representations of SO($4N$) of dimensions 2^{2N-1}. We can use two component Weyl spinors and rewrite (13) as follows

$$\{Q^i_\alpha, \bar{Q}_{\dot\beta}{}^j\} = \delta_{\alpha\dot\beta}\delta^{ij}, \qquad \{Q^i_\alpha, Q^j_\beta\} = 0 \qquad (\alpha, \beta = 1, 2) \tag{14}$$

Q^i_α, $\bar{Q}^i_\alpha$ satisfy the algebra of $2N$ fermionic creation and destruction operators. If we start from a Clifford vacuum Ω defined by the condition

$$Q^i_\alpha \Omega = 0 \qquad (\forall\ \alpha, i), \tag{15}$$

we build up the 2^{2N} states in the following way

$$\Omega,\ \bar{Q}^i_{\dot\alpha}\Omega,\ \bar{Q}^{i_1}_{\dot\alpha_{i_1}}\bar{Q}^{i_2}_{\dot\alpha_{i_2}}\Omega, \dots \bar{Q}^{i_1}_{\dot\alpha_{i_1}} \dots \bar{Q}^{i_n}_{\dot\alpha_{i_n}}\Omega, \dots \tag{16}$$

These states are classified by an instrinsic spin operator

$$W_k = \frac{1}{4}\,\sigma_k{}^{\alpha\dot\alpha}[Q^i_\alpha, \bar{Q}_{\dot\alpha}{}^i] \tag{17}$$

which belongs to the enveloping algebra of GLA. If we define the $2N$ component spinors

$$Q^a_\alpha = Q^i_\alpha \quad \text{for} \quad i = 1 \dots N, \qquad Q^a_\alpha = \bar{Q}^{\dot\alpha i} = \varepsilon^{\dot\alpha\dot\beta}\bar{Q}^i_{\dot\beta} \quad \text{for} \quad a = N+1 \dots 2N \tag{18}$$

then (13) and (14) become

$$\{Q^a_\alpha, Q^b_\beta\} = \varepsilon_{\alpha\beta}\Omega^{ab} \qquad (\alpha, \beta = 1, 2, \quad a, b = 1 \dots 2N)$$

and

$$\Omega^{ab} = -\Omega^{ba} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \qquad (Q^a_\alpha)^* = \Omega_{ab}\varepsilon^{\alpha\beta}Q^b_\beta \tag{19}$$

The algebra given in (19) has a manifest SU(2) $\otimes$ USp($2N$) covariance [25, 26]. The $4N$ dimensional vector representation of SO($4N$) remains irreducible in the group decomposition

$$\mathrm{SO}(4N) \to \mathrm{SU}(2) \otimes \mathrm{USp}(2N); \quad 4N \to (2, 2N) \tag{20}$$

The symplectic generators which classify the states in (16) are given by

$$A^a{}_b = T^a{}_b + T^b{}_a + i(T^a{}_b - T^b{}_a); \quad T^{ab} = \varepsilon^{\alpha\beta}[Q^a_\alpha, Q^b_\beta] \tag{21}$$

and they commute with the SU(2) generators (17). The two spinor representations of SO(4N) decompose into a sum of irreducible representations as follows

$$2^{2N} = (N+1, 1) + (N, 2N) + \cdots (N-K+1, [2N\times \cdots 2N]_k) \\ + \cdots (1, [2N\times \cdots 2N]_N) \tag{22}$$

$[2N\times \cdots 2N]_k$ is the k-fold antisymmetric traceless representation of USp(2N). The two irreducible representations of SO(4N), which correspond to the ± 1 eigenvalues of the γ^{4N+1} operator, separate the integral (bosons) and half-integral spin fermions in the decomposition (22). The SU(2) spin J, in the spinor representations of SO(4N), runs from $J = 0$ up to $J = N/2$. Let us consider, for definiteness, as an explicit example, the case $N = 2$. Then the Clifford algebra corresponds to the group SO(8) and the Q^i_α generators, which are vectors under SO(8), are embedded in the (2,4) irreducible representation of SU(2) $\otimes$ USp(4) $\sim$ SO(3) $\otimes$ SP(5) $\subset$ SO(8). From the Dynkin diagrams of D_4 [27] is it known that the three inequivalent eight-dimensional representations of SO(8): 8_v, 8, 8′ decompose into: (2,4), (2,4), (3,1) + (1,5) (or permutations) under SO(3) $\otimes$ SO(5). We conclude that, in the fundamental massive supermultiplet of $N = 2$ extended symmetry, the fermions and the bosons transform as (2,4) and (3,1) + (1,5) respectively under SU(2) $\otimes$ USp(4).

More general representations of the supersymmetry algebra can be constructed. We can relax the condition that the Clifford vacuum Ω is a singlet under the spin group generated by the Pauli-Lubanski-Bargmann operator and assume that it belongs to a non-trivial representation of the spin group and of the internal symmetry. Then a general irreducible massive representation of N extended supersymmetry will be classified by several Casimir operators, the superspin and the Casimir operator of the internal symmetry part [23, 28]. In superfield language it is easy to write down the Casimir operators in terms of covariant derivatives. In fact, if we replace, in all previous expressions, the Q^a_α operators with the covariant derivatives D^a_α we get operators which commute with the supersymmetry generators because $\{Q^a_\alpha, D^b_\beta\} = 0$. Then the Casimir operators of the algebra constructed out of the D^a_α can be used to classify the irreducible representations of N extended supersymmetry. In Table 1 some massive representations of N-extended supersymmetry are reported (up to $N = 5$). Massive representations of N-extended supersymmetry have dimensions

$$d = 2^{2N} \times d_\Omega \tag{23}$$

where d_Ω is the dimension of the Clifford vacuum. Massive representations can have smaller dimensions in the presence of central charges [29, 30]. We will give a few examples: in $N = 2n$ extended supersymmetry massive multiplets with central charges have a dimension 2^{2n+1} instead of 2^{4n}. The spin range is $J = 0$ up to $J = n/2$ instead of $J = 0$ up to $J = n$. These 2^{2n+1} states are a doublet of massive representations of

Table 1. *Some massive representations (without central charges)*

$J\rightarrow$ $N\downarrow$	$\frac{5}{2}$	2	$\frac{3}{2}$	1	$\frac{1}{2}$	0
1		1	1 2	1 2 1	1 2 1	2 1
2		1	1 4	1 4 $5 \oplus 1$	4 $5 \oplus 1$ 4	5 4 1
3		1	1 6	6 $14 \oplus 1$	14 $14 \oplus 6$	14 14
4		1	8	27	48	42
5	1	10	44	110	165	132

Table 2. *Some massive representations (with central charges). Complex representations*

$J\rightarrow$ $N\downarrow$	$\frac{3}{2}$	1	$\frac{1}{2}$	0
2	1	1 2	1 2 1	2 1
4	1	1 4	4 $5 \oplus 1$	5 4
6	1	6	14	14

n extended supersymmetry without central charges. The internal symmetry which classifies these states is $U(1) \otimes USp(2n)$ instead of $USp(4n)$. In Table 2 some massive representations with non-vanishing central charges are reported.

Let us turn to massless representations. In this case, we may choose $p^\mu = (1, 0, 0, 1)$ and the stability subalgebra written in terms of Weyl spinors becomes

$$\{Q^i_\alpha, \bar{Q}_{\dot\beta}{}^j\} = (1 + \sigma_3)_{\alpha\dot\beta}\delta^{ij}, \qquad \{Q^i_\alpha, Q^j_\beta\} = 0 \tag{24}$$

(24) implies that $Q^j_2 = 0$ and we have the Clifford algebra for N creation and destruction operators. If we set $Q^i_1 = Q_i$ we have (after the rescaling of Q^i by a factor $\sqrt{2}$)

$$\{Q^i, \bar{Q}^j\} = \delta^{ij}, \qquad \{Q^i, Q^j\} = 0 \tag{25}$$

If we define the real vector

$$Q^a = \left[\frac{Q^i + \bar{Q}^i}{\sqrt{2}}, \frac{i(Q^i - \bar{Q}^i)}{\sqrt{2}}\right] \quad (a = 1, \ldots, 2N)$$

(25) becomes the Clifford algebra of SO($2N$). The fundamental massless multiplet of N extended supersymmetry coincides with the spinor representation of SO($2N$) and has a dimension 2^N. We will be interested in the decomposition of SO($2N$) into U(N) corresponding to the embedding $\bar{N} + N = 2N$ of the fundamental SU(N) representation in the SO($2N$) vector representation. In fact, when $M = 0$ the USp($2N$) intrinsic group defined by (21) reduces to U(N) with generators

$$\tilde{T}^{ij} = T^{ij} + T^{ji} + i(T^{ij} - T^{ji}), \qquad T^{ij} = \frac{1}{4}[\bar{Q}^i, Q^j] \tag{26}$$

The U(1) part $T = 1/4[\bar{Q}^i, Q^i]$ is related to the intrinsic helicity $\Lambda = 1/2\bar{Q}^i Q^i (\Lambda\Omega = 0)$ by the following relation

$$T = \Lambda - \frac{N}{4} \tag{27}$$

If we consider the representation space spanned by $\Omega, \bar{Q}^i\Omega, \ldots \bar{Q}^{i_1} \ldots \bar{Q}^{i_k}\Omega, \ldots$ the spectrum of T runs from $-N/4$ up to $N/4$. If we define the Poincaré helicity as that operator Γ which transforms the $\bar{Q}^i$ opposite to T then the superhelicity is given by $\Gamma + T$ and is a Casimir invariant for a massless representation. If we want a multiplet to be PCT self-conjugate we must have that $\Gamma + T = 0$, i.e., $\Gamma\Omega = N/4\Omega$. In this case the multiplet contains a complete set of opposite helicity states $1/2(N/2 - k)$ $k = 0, \ldots N$ which belong to irreducible representations of SU(N): $[\bar{N} \times \bar{N} \times \cdots \times \bar{N}]_k$. If $\Gamma\Omega = \lambda\Omega$ with $\lambda \neq N/4$ then we must add to the multiplet with superhelicity $\lambda - N/4$ a PCT conjugate multiplet with opposite superhelicity $N/4 - \lambda$. This means that $\Gamma\Omega' = (N/2 - \lambda)\,\Omega'$. Note that a multiplet is never PCT self-conjugate if N is odd. We can enlarge the dimension of the Clifford vacuum by assuming that Ω transforms according to some irreducible representation R of the chiral group SU(N).

Table 3. *Massless representations with maximum helicity $\lambda = 1$*

$\lambda \rightarrow$ / $N \downarrow$	1	$\frac{1}{2}$	0
1	1	1	
2	1	2	$2 = 1 \oplus 1$
3	1	$3 \oplus 1$	$3 \oplus 3$
4	1	4	$3 \otimes 1 \oplus 1 \otimes 3$

In this case, a PCT self-conjugate representation of the supersymmetry algebra will be obtained by adding two representations with opposite superhelicities and Clifford vacua transforming according to the R and $\bar{R}$ representations of SU(N) respectively. Again, the doubling is not necessary if the superhelicity vanishes and if R is a self-conjugate representation of SU(N). The dimensionality of an arbitrary representation will be $2^N \times \dim R$ if the superhelicity vanishes or $2^{N+1} \times \dim R$ otherwise. From the previous considerations it follows that the minimum helicity range for a PCT self-conjugate massless multiplet is $\lambda = 0$ up to $\lambda = N/4$ (or $(N+1)/4$ for odd N). In Tables 3 and 4 massless multiplets with $\lambda_{\max} = 1$ and 2, singlets under SU(N), are shown. We note that, as far as the internal symmetry properties are concerned, the SO(N) subgroup of SU(N) is the maximal internal symmetry of the massless multiplets for which the representations are vector-like. When SO(N) is enlarged to SU(N) the massless multiplets are not invariant under parity. This is because a given SU(N) representation acts on a state of given chirality which has no partner of opposite chirality. However, it can happen that some states in a given multiplet belong to a self-conjugate representation of SU(N). In this case we can define a parity operation on these states. It is evident that a given irreducible PCT self-conjugate multiplet is never self-conjugate under the chiral SU(N) symmetry which rotates the generic state $\bar{Q}^{i_1} \ldots \bar{Q}^{i_k}\Omega$. However, we can add several massless multiplets in order to obtain only states which are "vector-like", i.e., states which belong to self-conjugate representations of SU(N). This is, for instance, the case when we decompose the states of a massive representation with respect to states of a given helicity [26]. The massive representation decomposes into a sum of irreducible massless representations. The states of a given helicity are classified by the subgroup SU(N) of USp($2N$) according to the decomposition $2N \to N + \bar{N}$ of the vector representation of USp($2N$) with respect to SU(N). In Table 5 we show such decompo-

Table 4. *Massless representations with maximum helicity $\lambda = 2$. SO(N) classification*

$\lambda \rightarrow$ $N \downarrow$	2	$\frac{3}{2}$	1	$\frac{1}{2}$	0
1	1	1			
2	1	2	1		
3	1	3	3	1	
4	1	4	$3 \otimes 1 \oplus 1 \otimes 3$	4	$2 = 1 \oplus 1$
5	1	5	10	$10 \oplus 1$	$5 \oplus 5$
6	1	6	$15 \oplus 1$	$20 \oplus 6$	$15 \oplus 15$
7	1	$7 \oplus 1$	$21 \oplus 7$	$35 \oplus 21$	$35 \oplus 35$
8	1	8	28	56	$35 \oplus 35$

sition for the massive multiplet with spin $J_{\max} = 2$ of $N = 4$ extended supersymmetry (fundamental representation).

From the previous analysis we find that the helicity range inside a massless representation is lower than in a massive representation. This is due to the fact that the "rest frame" Clifford algebra corresponds to SO($2N$) rather than SO($4N$). The massless fundamental representation of N-extended supersymmetry is nothing but the spinor representation of SO($2N$) which corresponds to the particular decomposition $2N \to N + \overline{N}$ of SO($2N$) into SU(N). It is clear that the previous discussion can be generalized to supersymmetries in any number of space-time dimensions. If, for instance, we go in ten dimensions [31] and we consider massless representations then the Clifford algebra is again SO(8), as for the $N = 2$ massive case in four dimensions. In this case, the two irreducible SO(8) representations are just the vector and the other spinorial representation. These two SO(8) representations describe a vector field and a Majorana-Weyl spinor in ten dimensions. If we go to 11 dimensions [32], we have the Clifford algebra of $N = 8$ massless multiplets or $N = 4$ massive multiplets in four dimensions.

We conclude this section by considering again some representations of the N-extended supersymmetry algebra with a non-vanishing central charge [29, 30]. We have seen that these representations are classified by USp(N) and have a spin range from 0 up to $N/4$ (N even). Their dimension is 2^{N+1} and they coincide, as far

Table 5. *Decomposition of the massive $N = 4$, $J_{\max} = 2$ representation into massless representations*

Helicity	Multiplicity	Irreducible massless multiplets				
$+2$	1 (1)	1				
$+\frac{3}{2}$	8 (8)	$\bar{4}$	4×1			
$+1$	28 (27 + 1)	6	$4 \times \bar{4}$	6×1		
$+\frac{1}{2}$	56 (48 + 8)	4	4×6	$6 \times \bar{4}$	$\bar{4} \times 1$	
0	70 (42 + 27 + 1)	1	4×4	6×6	$\bar{4} \times \bar{4}$	1
$-\frac{1}{2}$	56 (48 + 8)		4×1	6×4	$\bar{4} \times 6$	$\bar{4}$
-1	28 (27 + 1)			6×1	$\bar{4} \times 4$	6
$-\frac{3}{2}$	8 (8)				$\bar{4} \times 1$	4
-2	1 (1)					1

Between brackets: USp(8) representations.

as the helicity content is concerned, with the PCT self-conjugate massless representations of N-extended supersymmetry. These representations can be considered as a particular embedding of USp(N) into SU(N). The SU(N) N-dimensional vector representation remains irreducible when restricted to USp(N). If we apply the above analysis to $N = 4$ supersymmetric Yang-Mills theories we conclude that any Higgs mechanism which preserves supersymmetry must necessarily give rise to massive multiplets with non-vanishing central charges whose spin states are classified by USp(4). This is indeed what happens in spontaneously broken extended super Yang-Mills theories.

4. Extended supergravity and particle physics

Extended supergravities [33] are the gauge theories of graded Poincaré-Lie algebras in which an N-plet of Majorana spinor charges Q^i_α ($i = 1 \ldots N$) is introduced through the basic anticommutation relations given in (2). These theories, because of the existence of N-spinor generators, require the introduction of N spin 3/2 Rarita-Schwinger fields which carry an internal symmetry index i. The particle content of the supergravity multiplets of N extended supergravity can be read from Table 4. Because multiplets with helicity $\lambda \leqq 2$ exist only up to $N = 8$ supersymmetry generators, it follows that there is a very limited number of pure supergravity theories. In fact, when the only coupling constant is the gravitational constant, there are only seven possible theories with different particle content. However, it has recently been shown [34] that different field representations can give inequivalent quantum field theories, which coincide at the classical level.

The simplest extended supergravity theory is the $N = 2$ theory [35]. It gives a very elegant unification of the Maxwell theory with the Einstein Lagrangian. This theory comes very close to Einstein's dream of unifying photons and gravitons and does so by adding two gravitini to the ordinary Maxwell-Einstein theory. The two gravitini are chargeless, but they interact with the Maxwell field via a non-minima coupling dictated by local supersymmetry,

$$\frac{1}{2}\, ek\bar{\psi}_\mu{}^i \left(F^{\mu\nu} + \frac{1}{2}\,\gamma_5\tilde{F}^{\mu\nu}\right) \psi_\nu{}^j\varepsilon_{ij} \tag{28}$$

This theory is an interacting theory whose free part reduces to the sum of the Lagrangians given by (4) and (7). It is also possible [36] to introduce in this theory a charge coupling g for the gravitino doublet provided a "mass term" of strength g/k and a cosmological term of strength $(g/k^2)^2$ are also introduced. Extended supergravities have a very large number of symmetries. Among the local symmetries, beyond general co-ordinate, Lorentz and supersymmetry invariances, they also have gauge symmetries related to the vector fields of the supergravity multiplet. These latter symmetries, when only the gravitational constant is present, are related [37] to the gauging of the central charges which appear on the right-hand side of the basic anticommutator of two supersymmetry generators, as given by (2). Among the global symmetries they have a global SO(N) invariance which can be enlarged to U(N). This is done [38] by a combined chiral and Maxwell-duality invariance which generalizes the Maxwell-duality invariance of the Maxwell-Einstein system. The N left-handed gravitini $\psi^i_{\mu L}$ transform as SU(N) vectors while the $N(N-1)/2$ Maxwell field strengths $F^{ij}_{\mu\nu} = -F^{ji}_{\mu\nu}$ and their dual $\tilde{F}_{\mu\nu} = 1/2\varepsilon_{\mu\nu}{}^{\varrho\sigma}F_{\varrho\sigma}$ transform as the two-fold antisymmetric representation of SU(N). This combined chirality-duality invariance, together with the other local symmetries, has been proven crucial for the one and two-loop finiteness of the quantum corrections of extended supergravity theories [12]. It is important to point out that the combined chirality-duality invariance induces new contact terms of gravitational strength for the spin-3/2 fields which do not reduce simply to torsion as in $N = 1$ supergravity. This fact shows that local supersymmetry has a much richer geometrical structure than the Einstein-Cartan theory of gravitation.

The global symmetries of extended supergravity have a much more interesting structure when $N > 4$. Under these circumstances, scalar fields appear in the supergravity multiplets (see Table 4) which now contain all possible helicity states from 0 up to two. It was shown [39], in the construction of $N = 4$ supergravity, that the two real scalar fields are responsible for an additional global SU(1,1) invariance of the field equations. This symmetry combines, in a non-trivial way, chiral transformations on the spinors, chirality transformations of the vector field strengths and projective transformations on the scalars. This non-compact global transformation extended the full global symmetry of $N = 4$ supergravity to SU(4) $\otimes$ SU(1,1). Some time ago, Cremmer and Julia [40] further generalized these additional symmetries of $N = 4$ supergravities by showing that all these theories have, in fact, a local symmetry $H = \mathrm{U}(N)$ (SU(8) for $N = 8$) and a global non-compact symmetry G. Furthermore, the scalar fields of the supergravity multiplet, which belong to the four-fold antisymmetric representation of the classifying group SU(N), parametrize the coset space $G/\tilde{H}$. H is related to G in the sense that it is isomorphic to its maximal compact subgroup $\tilde{H} \approx H$. This is the very reason for the absence of ghost states in spite of the fact that the over-all symmetry of the equations of motion is a non-compact one. In $N = 4$ supergravity $G/\tilde{H} = \mathrm{SU}(1,1)/\mathrm{U}(1)$ is a two-dimensional manifold and it corresponds to the two scalar modes of the theory. Let us jump immediately to the maximally extended $N = 8$ theory. In this case, $G = E_7$ and $H = \mathrm{SU}(8)$ [40]. The 70-scalar fields parametrize the homogeneous space G/H of dimension $133 - 63 = 70$. The SU(8) group which classifies the states is neither H nor $\tilde{H}$, but rather its direct sum $H \otimes \tilde{H}$. The local SU(8) group of $N = 8$ extended supergravity has a connection $Q_\mu{}^A{}_B$ which is an auxiliary, non-propagating field whose linearized part is a bi-linear expression in terms of the 70-scalar fields of the basic $N = 8$ supergravity multiplet. Cremmer and Julia [40] suggested that these SU(8) gauge vector potentials could become dynamical, namely their propagator could develop a zero mass pole, in analogy with a similar phenomenon which actually takes place in CP^{n-1} bi-dimensional non-linear σ models in the $1/n$ expansion [41]. This assumption is nothing more than the statement that the elementary fields of the $N = 8$ supergravity Lagrangian bind to form a multiplet which contains the adjoint representation of SU(8). This observation may be relevant in order to make contact between supergravity and current particle phenomenology. In fact, previous attempts which identify the elementary particles of "low energy" physics with the partners of the graviton in the massless supergravity multiplet failed essentially because N-extended supergravity can, at the most, accommodate an SO(8) Yang-Mills theory [42]. The elementary vector particles belong precisely to the adjoint representation of SO(8). This group, in spite of the fact that it has rank 4, does not contain SU(3) $\otimes$ SU(2) $\otimes$ U(1), the minimal gauge group of low-energy Physics. Even if we try to identify only the exact (vector-like) symmetry SU(3) $\otimes$ U(1), as a subgroup of SO(8), too many observed states of quarks and leptons are missing. However, if we abandon the constituent picture of the elementary fields of $N = 8$ supergravity and try instead to interpret them as "preons" of composite states which look elementary at present energies (quarks, leptons, vector bosons), then the local SU(8) chiral symmetry may be sufficient to accommodate all observed states. In a very interesting work, Ellis, Gaillard, Maiani and Zumino [43] recently made a group-theoretical analysis of the particle content of the SU(8) composite massless multiplet which contains the $\lambda = \pm 1$ helicity states in the adjoint representation of SU(8) and concluded that this multiplet is large enough in the spin 1/2 and 0 sector to include all observed quark and lepton degrees of freedom as well as the Higgs particles needed for subsequent symmetry breakings from the Planck mass down to the GeV region:

$$\mathrm{SU}(8) \to \mathrm{SU}(5) \to \mathrm{SU}(3) \otimes \mathrm{SU}(2)_{\mathrm{left}} \otimes \mathrm{U}(1) \to \mathrm{SU}(3) \otimes \mathrm{U}(1).$$

It is remarkable that SU(8) appears to be large enough not only to give the GUT SU(5) gauge group in a unique fashion, but also to give a family generation group due to the splitting of SU(8) into SU(5) $\otimes$ $\mathrm{SU}(3)_{\mathrm{families}}$. In order to construct the PCT self-conjugate massless multiplet which contains the vector particles in the adjoint representation of SU(8), Ellis et al. [43] have used the fact that, in any supersymmetric theory with $N \leqq 4$, there is a multiplet of currents which have the same on mass-shell degrees of freedom as an N extended massive multiplet [44]. This multiplet has the spin content

$$2, \left(\frac{3}{2}\right)^a, \ldots \tag{29}$$

where the index a runs from 1 up to $2N$ and the lower spin states belong to antisymmetric representations of USp($2N$) according to the classification given in Section 3. In the limit of zero mass, this multiplet decomposes in several massless multiplets which are classified by SU(8) representations according to the decomposition $2N \to N + \overline{N}$ of USp($2N$) into SU(N). For $N = 4$ we have, for instance, the decomposition given in Table 5. The massless multiplet which contains the adjoint representation of SU(4) is the one starting with a $+3/2$ helicity state in the vector representation of SU(4). This is true for all $N \leqq 4$. Ellis et al. [43] assumed that for $N > 4$ the relevant multiplet is also given by

$$\left(\frac{3}{2}\right)^A (1)^A{}_B \left(\frac{1}{2}\right)^A{}_{[BC]} \cdots \left(\frac{3-N}{2}\right)^A \tag{30}$$

to which one must add the PCT conjugate states

$$\left(\frac{N-3}{2}\right)_A \left(\frac{N-4}{2}\right)_{AB} \left(\frac{N-5}{2}\right)_{A[BC]} \cdots \left(\frac{-3}{2}\right)_A \tag{31}$$

For $N = 8$ the over-all helicity content for left-handed particle states is

$$\begin{aligned}
\lambda &= \mp\frac{5}{2}\colon\ 8 \\
\lambda &= \mp 2\colon\ 28 + 36 \\
\lambda &= \mp\frac{3}{2}\colon\ \bar{8} + 56 + 168 \\
\lambda &= \mp 1\colon\ 63 + 1 + 70 + 378 \\
\lambda &= \mp\frac{1}{2}\colon\ 8 + 216 + \overline{56} + \overline{504} \\
\lambda &= 0\colon\ 28 + 420 + \overline{28} + \overline{420}
\end{aligned} \tag{32}$$

The above authors made the drastic assumption of neglecting SU(8) trace representations in (32). For the spin 1/2 these representations are the 8 and $\overline{56}$ and for the scalar sector the $28(\overline{28})$. Under these circumstances, they conclude that the maximal unbroken subgroup of SU(8) below the Planck scale 10^{19} GeV is SU(5). The maximal anomaly free subset of spin 1/2 left-handed states which can be constructed out of 216 and $\overline{504}$ representations and which is vector-like under $SU(3)_{\text{colour}} \otimes U(1)_{\text{em}}$ contains exactly three families of $(10 + \bar{5})$ SU(5)-representations plus a set of self-conjugate representations which may acquire a big SU(5) invariant mass of order $\sim 10^{15}$ GeV. Some possible alternatives can be envisaged in the approach considered by Ellis et al. One should look to anomaly-free representations for spin 1/2 left-handed fermions without disregarding SU(8) trace representations. Another possibility which seems hard to fulfil is to restore conventional symmetry breaking by considering more multiplets other than those given by (30) and (31). It is evident that if one adds enough massless supermultiplets to (32) one can, in fact, give a supersymmetric mass to every state by enlarging SU(8) to USp(16). It is easy to see that this is achieved by a massive $N = 8$ multiplet with $J_{\max} = 6$. This is too much, but one could explore intermediate situations which fulfil the constraints imposed by low energy phenomenology. It remains to be seen if this programme can have any solution.

References

[1] J. C. Pati and A. Salam, Phys. Rev. Lett. **31** (1973) 661; Phys. Rev. **D 8** (1973) 1240; H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32** (1974) 438.

[2] For a review see: J. Ellis, Proceedings of the EPS International Conference on High Energy Physics, Geneva, **2** (1979) 940.

[3] See for instance D. V. Nanopoulos, preprint CERN TH. 2896 (1980), lectures given at the XV Rencontre de Moriond, les Arcs, March 1980; preprint CERN TH. 2866 (1980) to appear in the Proceedings of the Europhysics Study Conference "Unification of the Fundamental Interactions", Erice (March 1980).

[4] See for instance: R. Barbieri, preprint CERN TH. 2850 (1980), to appear in the Proceedings of the Europhysics Study Conference "Unification of the Fundamental Interactions", Erice, (March 1980).

[5] D. Z. Freedman, P. van Nieuwenhuizen and S. Ferrara, Phys. Rev. **D 13** (1976) 3214.

[6] S. Deser and B. Zumino, Phys. Lett. **62 B** (1976) 335.

[7] Y. A. Gol'fand and E. P. Likhtam, JEPT Lett. **13** (1971) 323; D. V. Volkov and V. P. Akulov, Phys. Lett. **46B** (1973) 109; J. Wess and B. Zumino, Nucl. Phys. **B 70** (1974) 39.

[8] For a review see: P. Fayet and S. Ferrara, Phys. Rep. **32** (1977) 251.

[9] R. Haag, J. T. Lopuszanski and M. Sohnius, Nucl. Phys. **B 88** (1975) 257.

[10] For a review see for instance: D. Z. Freedman in Recent Development in Gravitation, edited by M. Levy and S. Deser (Plenum Publishing Corporation, 1979) page 549.

[11] G. Velo and D. Zwanziger, Phys. Rev. **186** (1969) 1337 and **188** (1969) 2218.

[12] P. van Nieuwenhuizen and M. T. Grisaru in "Deeper Pathways in High Energy Physics" Coral Gables (ed A. Perlmutter and L. F. Scott, 1977) page 233.

[13] B. de Wit and D. Z. Freedman, Nucl. Phys. **B 130** (1977) 105; E. Cremmer an B. Julia, Phys. Lett. **80B** (1978) 48; Nucl. Phys. **B 159** (1979) 141.

[14] M. J. Duff, ICTP. 79/80/38 talk presented at the Second Oxford Quantum Gravity Conference, April 1980.

[15] A. Chamseddine and P. C. West, Nucl. Phys. **B 129** (1977) 34.

[16] S. W. MacDowell and F. Mansouri, Phys. Rev. Lett. **38** (1977) 139.

[17] J. Wess and B. Zumino, Phys. Lett. **66B** (1977) 361.

[18] Y. Ne'eman and T. Regge, Riv. Nuovo Cimento, **1** (1978) 1; A. D'Adda, R. D'Auria, P. Fré and T. Regge, Riv. Nuovo Cimento, **3** (1980) 71.

[19] L. Brink, M. Gell-Mann, P. Ramond and J. H. Schwarz, Phys. Lett. **76B** (1978) 417.

[20] V. Ogievetsky and E. Sokatchev, Phys. Lett. **79B** (1978) 222.

[21] W. Siegel and J. Gates, Nucl. Phys. **B 147** (1979) 77.

[22] S. Ferrara, F. Gliozzi, J. Scherk and P. van Nieuwenhuizen, Nucl. Phys. **B 117** (1976) 333.

[23] A. Salam and J. Strathdee, Nucl. Phys. **B 76** (1974) 4778; Nucl. Phys. **B 84** (1975) 127.

[24] W. Nahm, Nucl. Phys. **B 135** (1978) 149.

[25] S. Ferrara and B. Zumino, Phys. Lett. **86B** (1979) 279.

[26] S. Ferrara, LNF-80/31(P), invited talk given at the Second Oxford Quantum Gravity Conference, April 1980.

[27] E. B. Dynkin, American Mathematical society translations, Series 2 vol 6 (1957).

[28] J. G. Taylor, Nucl. Phys. **B 169** (1980) 484.

[29] M. Sohnius, Nucl. Phys. **B 138** (1978) 109.

[30] P. Fayet, Nucl. Phys. **B 149** (1979) 137; preprint LPTENS 80/7 (1980). Lectures given at the XVII Winter School of Theoretical Physics, Karpacz (Poland), February 1980.

[31] F. Gliozzi, J. Scherk and D. Olive, Nucl. Phys. **B 122** (1977) 253; L. Brink, J. H. Schwarz and J. Scherk, Nucl. Phys. **B 121** (1977) 77.

[32] E. Cremmer, B. Julia and J. Scherk, Phys. Lett. **84B** (1979) 93.

[33] For a recent review, see, P. van Nieuwenhuizen, Phys. Rep. to appear.

[34] M. J. Duff and P. van Nieuwenhuizen, Phys. Lett. **94B** (1980) 179;
A. Aurilia, H. Nicolai and P. K. Townsend, preprint CERN TH. 2875 (1980);
V. Ogievetsky and E. Sokatchev, Dubna preprint E2-80-19 (1980).
[35] S. Ferrara and P. van Nieuwenhuizen, Phys. Rev. Lett. **37** (1976) 1669.
[36] D. T. Freedman and A. Das, Nucl. Phys. **B 120** (1977) 221;
E. S. Fradkin and M. A. Vasiliev, Lebedev Institute preprint (1977).
[37] S. Ferrara, J. Scherk and B. Zumino, Phys. Lett. **66B** (1977) 35.
[38] S. Ferrara, J. Scherk and B. Zumino, Nucl. Phys. **B 121** (1977) 393.
[39] E. Cremmer, H. Scherk and S. Ferrara, Phys. Lett. **74B** (1978) 61.
[40] E. Cremmer and B. Julia as in Ref. [13]; see also E. Cremmer preprint LPTENS 80/9 (1980) to appear in the Proceedings of the Europhysics Study Conference "Unification of the Fundamental Interactions", Erice, (March 1980).
[41] A. d'Adda, P. Di Vecchia and M. Luscher, Nucl. Phys. **B 146** (1978) 63; Nucl. Phys. **B 152** (1979) 125;
E. Witten, Nucl. Phys. **B 149** (1979) 285.
[42] M. Gell-Mann, talk given at the 1977 Washington Meeting of the American Physical Society.
[43] J. Ellis, M. K. Gaillard, L. Maiani and B. Zumino, LAPP preprint TH-15/CERN preprint TH-2481 (1980) to appear in the Proceedings of the Europhysics Study Conference on "Unification of the Fundamental Interactions", Erice, (March 1980).
J. Ellis, M. K. Gaillard and B. Zumino, Phys. Lett. **94B** (1980) 343;
B. Zumino, FERMILAB-Conf-80/74-THY — Talk presented at the XX International Conference on High-Energy Physics, July 1980, Madison Wisconsin.
[44] S. Ferrara and B. Zumino, Nucl. Phys. **B 87** (1975) 207.

Address: Prof. Dr. S. Ferrara
Theory Division
CERN
1211 Geneva 23
Switzerland

The Einstein-Cartan Theory

A. Trautman (Warsaw)

Summary

(Full text of this lecture will be published elsewhere)

In 1923 Elie Cartan proposed to modify the Einstein theory of gravitation by allowing space-time to have torsion and relating it to the density of intrinsic angular momentum of a continuous medium. Cartan's idea did not attract any attention at that time. This may be due, on the one hand, to the fact that Cartan's paper had appeared before the spin of the electron was discovered, and, on the other, to Einstein's fascination with the problem of unifying gravity with electromagnetism.

The idea of connecting torsion to spin became alive again around 1960, mainly thanks to the work of D. W. Sciama and T. W. B. Kibble. There was considerable activity on this problem from 1965 to 1975.

There is no "logical" or experimental, compelling need to modify Einstein's theory, but one can advance good heuristic arguments in favour of the Cartan idea:

(i) The geometrical independence of the metric g and linear connection Γ leads to the idea of treating these quantities as independent variables in the sense of a principle of least action. If g and Γ are assumed to be compatible, then the freedom in the choice of Γ reduces to that of the torsion tensor Q.

(ii) According to relativistic quantum theory, the Poincaré — or the inhomogeneous Lorentz group — is physically more significant than the Lorentz group itself. The Poincaré group has two fundamental invariants: mass and spin. The first among them is related to translations and to energy-momentum. In Einstein's theory, the density of energy-momentum is source of curvature whereas spin has no such direct dynamical significance. In a sense, the Einstein-Cartan theory restores — to some extent — the symmetry between mass and spin. It introduces also an unexpected "duality": via Noether's theorem, energy-momentum is generated by translations whereas Einstein's equation relates it to curvature, which is responsible for rotations of vectors undergoing parallel transport. Conversely, spin is generated by rotations, but torsion induces translations in the tangent spaces to a manifold ("Cartan displacement"). This duality can be traced to the fact that the Einstein-Cartan Lagrangian is linear in curvature, an assumption criticized by C. N. Yang. Recently, F. W. Hehl, Y. Ne'eman, N. Straumann and their coworkers have studied a theory of gravitation based on a Lagrangian quadratic in both curvature and torsion. It is clear however, that there are no compelling reasons to abandon the linear Lagrangian.

(iii) There is an interesting analogy between the description of magnetic moments in electrodynamics and spin in the theory of gravitation. In a phenomenological

description of electromagnetism, the external magnetic field produced by a ferromagnet may be obtained in at least three ways: by considering a surface current equivalent to the actual distribution of microscopic currents and magnetic moments, by replacing the latter by a volume distribution of "Ampère currents", or, finally, by introducing a smooth field of the magnetization vector. In the Einstein theory, there are analogues for the first two descriptions, whereas the Einstein-Cartan theory provides the third.

The Einstein-Cartan theory assumes, as a model of spacetime, a four-dimensional manifold with a linear connection Γ compatible with a metric tensor g. The gravitational part of the Lagrangian, $\sqrt{-g}\,R$, is formed from the curvature tensor of Γ. This prescription leaves no room for new arbitrary constants. The left hand sides of the field equations are obtained by varying this Lagrangian with respect to g and Q. Variation with respect to g may be replaced by that relative to the field of frames (tetrads). The sources of the gravitational field are described by expressions resulting either from phenomenology or by varying an action integral obtained by applying the principle of minimal gravitational coupling to a special-relativistic Lagrangian. There are subtleties concerning the Maxwell and other gauge fields.

The Einstein-Cartan field equations are

$$R_{\mu\nu} - \frac{1}{2}\,g_{\mu\nu}R = \frac{8\pi G}{c^4}\,t_{\mu\nu} \tag{E}$$

$$Q^{\varrho}_{\mu\nu} - \delta^{\varrho}_{\mu}Q^{\sigma}_{\sigma\nu} - \delta^{\varrho}_{\nu}Q^{\sigma}_{\mu\sigma} = \frac{8\pi G}{c^3}\,s^{\varrho}_{\mu\nu} \tag{C}$$

The Cartan equation (C) is trivial in the sense that if the spin density vanishes, $s^{\varrho}_{\mu\nu} = 0$, then so does torsion, $Q^{\varrho}_{\mu\nu} = 0$. Quite independently of this, torsion is topologically trivial: any linear connection can be deformed into a connection without torsion.

The Bianchi identities for $R^{\mu}{}_{\nu\varrho\sigma}$ and $Q^{\varrho}_{\mu\nu}$ give two sets of constraints on the sources. One of them may be symbolically written as

$$\nabla t = R \cdot s + Q \cdot t \sim R \cdot Q$$

Without good reason, Cartan required $\nabla t = 0$ and was led to the algebraic constraint $R \cdot Q = 0$.

F. W. Hehl has shown that, by solving (C) for Q, one can reduce the system (E) — (C) to an equation with the Einstein tensor, built from g, on the left and an effective energy-momentum tensor,

$$T_{\text{eff}} = t + \operatorname{div} s + s^2 \tag{H}$$

on the right. The term quadratic in spin seems to be the only essential difference between the Einstein-Cartan and Einstein theories. Similar terms have also been obtained by B. M. Barker and R. F. O'Connell from Gupta's quantum theory of gravity.

On the basis of (H) one can argue that the Einstein-Cartan theory may be physically relevant only when the density of energy is of the same order of magnitude as the spin density squared. For matter consisting of particles of mass m and spin $\hbar/2$, this will occur at densities of order $m^2c^4/G\hbar^2$. For nucleons, the density in question is 10^{54} g/cm^3, much less than the Planck density. Putting the same result in a slightly different way, one can say that a particle should have a radius of order

$$\left(\frac{Gm^2}{\hbar c}\right)^{1/3} \frac{\hbar}{mc}$$

for gravitational effects of spin to be comparable to those of mass.

The Einstein-Cartan theory is viable, but differs so slightly from Einstein's theory that it may take a very long time to confirm it or disprove.

Address: Prof. Dr. A. Trautman
Uniwersytet Warszawski
Wydzial Fizyki
Instytut Fizyki Teoretycznej
ul. Hoża 69
00-681 Warszawa
Poland

The Positive Energy Program[1])

D. R. Brill (College Park) and P.-S. Jang (Syracuse, N.Y.)

1. Introduction: Description of the Problem

Some months ago I had the pleasure of meeting Konrad Wachsmann, the architect of Einstein's summer house near Potsdam. One of the things he said comes to mind: "I am a marked man. Everyone thinks of me in terms of Einstein's house, but that is something I did a long time ago when I was very young." Well, I am beginning to feel the same way about Positive Energy: it is something I worked on a long time ago, and the only recent paper I wrote on it is a review together with Pong Soo Jang [1]. Immediately after finishing that review there was a flurry of activity in the field, no doubt designed to outdate our review as quickly as possible. So it is now appropriate to assess the progress made since GR 8.

Thanks to the work of Jang, Schoen and Yau, and others, the positivity theorem has now been firmly established. Important further progress is possible for at least two reasons: (a) the theorem to be proved is becoming more ambitious, (b) the present proof involves many complicated steps and should be capable of simplification. So the positive energy program may still be an interesting topic three years hence at GR 10.

I confine attention in this talk to the "ADM" energy defined on a spacelike hypersurface. Its relation to other energy quantities, such as the Bondi energy, is another topic on which significant research is just beginning today [2]. We have already heard in other lectures of this conference (e.g. in that of Marsden) that unique expressions for energy and momentum on spacelike surfaces can be given if the geometry is asymptotically flat:

$$
\begin{aligned}
m = P^0 &= \lim_{r\to\infty} (1/16\pi) \int (\partial_j g_{ij} - \partial_i g_{jj})\, dS_i \\
P^i &= \lim_{r\to\infty} (1/8\pi) \int (K^i{}_j - \delta^i{}_j K)\, dS_j
\end{aligned}
\qquad (1)
$$

By "asymptotically flat" we mean roughly

$$
\begin{aligned}
g_{ij} &= \delta_{ij} + 0(1/r), \\
K_{ij} &= 0(1/r^2),
\end{aligned}
$$

recognizing of course that more precise definitions, e.g. in terms of weighted Sobolev spaces, are available and necessary for the rigorous proofs [3]. The requirement on

[1]) Supported in part by NSF grant PHY 79-06940

g_{ij} demands an asymptotic behavior, sufficiently close to the Schwarzschild spacelike geometry of mass m, that total energy is well defined; similarly, the fall-off requirement on the extrinsic curvature (second fundamental form) K_{ij} assures that the total momentum is well defined.

By the positive energy property we mean that P^μ is timelike and future pointing on any asymptotically flat slice of spacetime, and that it vanishes only if spacetime is flat. In particular, then, the total mass-energy m is non-negative, and vanishes only for flat spacetime.

It has been customary to prove the second, more particular statement, and then to appeal to the Lorentz transformation properties of P^μ to establish that it is timelike (since $P^0 > 0$ in all frames). However, the positive energy property makes sense for a single spacelike surface, and it would be appropriate to have a proof which does not involve more than one such surface. Only very recently has the interesting work of Christodoulou and O'Murchadha [4] given an indication that asymptotically flat Cauchy surfaces will generally be accompanied by other, boosted surfaces; but at present the boost is not known to reach arbitrarily large values.

The general idea for establishing the positive energy property is simple enough: Use the constraints,

$$\begin{aligned} &R - K^{ij}K_{ij} + K^2 - 2\varrho = 0 \\ &D_j(K^{ij} - Kg^{ij}) - J^i = 0 \end{aligned} \tag{2}$$

to express the energy in terms of interior gravitational fields and matter; if matter has positive energy, i.e. satisfies "energy dominance",

$$\varrho \geqq (J_i J^i)^{1/2} \tag{3}$$

then positivity of total energy should follow. The analogous situation in electrodynamics illustrates the procedure: the total charge is defined as an asymptotic surface integral, $Q = \int \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{S}$. By using the constraint, $\operatorname{div} \boldsymbol{E} = \varrho$, this can be expressed in terms of the interior charge density, $Q = \int \varrho \, \mathrm{d}V$; and if there were a positivity property of ϱ, the corresponding property of Q would immediately follow. The analogous steps in the case of gravity lead to an expression of the suggestive form

$$16\pi m = \int (g\, \partial g\, \partial g + KK - K^2 + 2\varrho) \, \mathrm{d}V .$$

The first term, quadratic in derivatives of the metric, could be considered a gravitational potential energy, the second term a gravitational kinetic energy, and the last term the matter energy. However, the "potential" term has no particular positivity property, and the "kinetic" term is positive only if $K = 0$, i.e. if the spacelike surface happens to be maximal.

2. Solution for the Case of Weak Fields

The only situation where the positive energy program can be carried out completely and straightforwardly is the case of weak fields. Here the metric and extrinsic curvature can be split up covariantly but nonlocally into (flat) background, transverse-traceless parts, and remainder:

$$g_{ij} = \delta_{ij} + \delta g_{ij}^{TT} + \delta \bar{g}_{ij}$$

$$\pi^{ij} := g^{1/2}(K^{ij} - g^{ij}K) = \delta\pi^{ijTT} + \delta\bar{\pi}^{ij}$$

where transverse and traceless means, e.g. for $\delta\pi^{ij}$,

$$D_j \delta\pi^{ijTT} = 0, \qquad \delta_{ij}\delta\pi^{ijTT} = 0.$$

By expressing P^0 in terms of this decomposition one can show that its first order deviation from the flat-space zero value vanishes, and that the second order expression is

$$8\pi\delta^2 P^0 = \int \left[\frac{1}{4}\,(\nabla\delta g_{ij}^{TT})^2 + (\delta\pi^{ijTT})^2 + \delta^2\varrho\right] \mathrm{d}^3x$$

Similarly the momentum has the second-order expression

$$8\pi\delta^2 P^i = \int [\partial^i \delta g_{kl}^{TT} \delta\pi^{klTT} + \delta^2 J^i]\, \mathrm{d}^3x$$

Therefore the desired timelikeness of $\delta^2 P^\mu$,

$$|\delta^2 P^i| \leqq \delta^2 P^0$$

follows manifestily from local energy dominance, $|\delta^2 J^i| < \delta^2\varrho$. Furthermore, if $\delta^2 P^0 = 0$ then we have

$$\partial_i \delta g_{jk}^{TT} = 0, \qquad \delta\pi^{ijTT} = 0$$

hence

$$\delta R_{ij} = 0, \qquad \partial_{[i}\delta K_{j]k} = 0,$$

which is the first order form of the Gauss-Codazzi equations for flat space-time, and which assures us that the initial surface is embedded in flat space-time.

Of course, these lowest order results do not mean anything directly for the exact theory. To establish a connection one needs: (1) a correspondence between solutions of the linearized field equations and solutions of the exact theory in a finite neighborhood of flat space-time. This is provided by the "linearization stability" of flat space-time, as explained e.g. by Marsden in his lecture; and (2) an estimate of the "remainder" between the exact energy-momentum and its quadratic approximation $\delta^2 P^\mu$, showing that this remainder is sufficiently small in some finite "local" neighborhood of flat space-time. This estimate was given by Choquet-Bruhat and Marsden [5] in their solution of the "local" positive energy problem.

3. Reduction to the Sourceless, Time-Symmetric Case

For the rest of this talk I shall confine attention to the proof of the positivity of *energy* in the general case. This proof proceeds in a number of steps. The first takes care of the "kinetic" part of the geometry, by reducing the problem to the sourceless, time-symmetric case, where the extrinsic curvature and the matter is entirely absent. This is easiest if the original data $\bar{g}$, $\bar{K}$ are given on a maximal surface, i.e. if $\operatorname{tr} \bar{K} = 0$. In this case the constraints imply $\bar{R} = R(\bar{g}) \geqq 0$, and we can apply a theorem of O'Murchadha and York [6]:

Every asymptotically flat metric $\bar{g}$ with $\bar{R} \geqq 0$, $\bar{R} \neq 0$ has greater P^0 than the unique conformally equivalent metric g which has $R = 0$ (i.e. which satisfies the time-symmetric, sourcefree constraints).

To prove this, let $\bar{g} = \theta^4 g$ and choose the scalar function θ such that $R(g) = 0$. One can show that such a function θ exists in the assumed case $R \geqq 0$. One then finds

$$\bar{R} = R\theta^{-4} - 8\theta^{-5}\Delta\theta = -8\theta^{-5}\Delta\theta$$

Also we have

$$P^0[\bar{g}] = -4 \int \theta_{,i} d^2S_i + P^0[g]$$

To finish the proof, we combine these equations to evaluate the energy difference,

$$P^0[\bar{g}] - P^0[g] = -4 \int \Delta\theta \,\mathrm{d}V = (1/2) \int \bar{R}\theta^5 \,\mathrm{d}V > 0.$$

This theorem was generalized by Schoen and Yau [7] to include the case of general initial surface, $K \neq 0$. In this case, define new initial data,

$$\bar{\bar{g}}_{ij} = \bar{g}_{ij} + \mathrm{D}_i w \mathrm{D}_j w$$

$$\bar{\bar{K}}^{ij} = \bar{K}^{ij} - (\mathrm{D}^i\mathrm{D}^j w)\left(1 + (\nabla w)^2\right)^{-1/2}$$

where w is a scalar function with $|\nabla w| = 0(1/r)$ satisfying Jang's equation [8]

$$\frac{\Delta w}{\left(1 + (\nabla w)^2\right)^{1/2}} - \frac{D^i w D^j w D_i D_j w}{\left(1 + (\nabla w)^2\right)^{3/2}} = \bar{K} - \frac{\bar{K}_{ij} D^i w D^j w}{1 + (\nabla w)^2}$$

Note that these new initial data have $\operatorname{tr} \bar{\bar{K}} = \bar{\bar{g}}_{ij}\bar{\bar{K}}^{ij} = 0$, and that the asymptotic dependence of $\bar{g}$ and $\bar{\bar{g}}$ is the same to $0(1/r)$, so the surface integral which defines P^0 is the same for both. Then make another, conformal change of the metric, $g = \Phi^4 \bar{\bar{g}}$, so that $R[g] = 0$. Schoen and Yau have proved existence of the functions w and Φ with the required properties. By expressing the constraints in terms of $\bar{\bar{g}}$, $\bar{\bar{K}}$, they can then again show that

$$P^0[\bar{g}] - P^0[g] > 0.$$

In this fashion the proof of positive energy in the general case is reduced to the proof that sourceless, time-symmetric gravitational waves have positive total energy. In some special cases — such as the case of axially symmetric waves [9] — this can be shown by converting the energy expression to a volume integral over a manifestly positive integrand. No such procedure is known in the general case. Instead one relies on cancellations between positive and negative contributions to the integrand. Theorems of global differential geometry, such as the Gauss-Bonnet theorem, show that such cancellations can occur in quite general situations. The Gauss-Bonnet theorem indeed plays a major role in the present methods of proof. Since this theorem has no convenient 3-dimensional form, these proofs use a further, 2 + 1 decomposition of the 3-dimensional spacelike geometry.

4. Proof of Schoen and Yau

We start with a metric $\bar{g}$ with $R[\bar{g}] = 0$ and $\bar{g}$ asymptotically flat. This means

$$\bar{g}_{ij} = \delta_{ij} + p_{ij} \quad \text{with} \quad p_{ij} = 0(1/r), \;\; p_{ij,k} = 0(1/r^2), \;\; p_{ij,kl} = 0(1/r^4);$$

but, as emphasized by York [10], $\bar{g}_{ij}$ does not need to be asymptotically Schwarzschildian in the sense

$$\bar{g}_{ij} = (1 + 2m/r)\, \delta_{ij} + 0(1/r^2).$$

(As a counterexample, York cites slices of the Schwarzschild geometry which are asymptotically boosted compared to the standard t = const. slices.) Therefore the aim of the first part of the Schoen-Yau proof [11] is to reduce the general asymptotically flat case to the asymptotically Schwarzschildian case, by showing that a change to asymptotically Schwarzschildian form can be made without changing the total energy appreciably:

Let $\bar{g}_{ij} = (1 + m/2r)^4\, \delta_{ij} + \tilde{g}_{ij}$, where m is the mass-energy of $\bar{g}$, $m = P^0[\bar{g}]$, so that $P^0[\tilde{g}] = 0$. Use $\tilde{g}$ to interpolate smoothly between $\bar{g}_{ij}$ and $(1 + m/2r)^4\, \delta_{ij}$ in a shell of radii σ and 2σ. The interpolating metric $\underset{\sigma}{\bar{\bar{g}}}$ depends on the arbitrarily chosen radius σ (and on the exact form of the smooth monotonically increasing function $\zeta(r)$ with $\zeta = 0$ for $r < \sigma$, $\zeta = 1$ for $r > 2\sigma$):

$$\underset{\sigma}{\bar{\bar{g}}} = \bar{g} - \underset{\sigma}{\zeta(r)}\, g.$$

Thus

$$\underset{\sigma}{\bar{\bar{g}}} = \bar{g} \quad \text{for} \quad r < \sigma \quad \text{and} \quad \underset{\sigma}{\bar{\bar{g}}} = (1 + 2m/r)^4\, \delta \quad \text{for } r > 2\sigma.$$

Within the shell we have $R[\underset{\sigma}{\bar{\bar{g}}}] \neq 0$, but by choosing σ large enough, this curvature is sufficiently small that it can be changed back to zero by a conformal transformation,

$$\underset{\sigma}{g} = \Phi^4 \underset{\sigma}{\bar{\bar{g}}}, \qquad R[\underset{\sigma}{g}] = 0.$$

One can then show that

$$P^0[\underset{\sigma}{g}] = P^0[\bar{g}] + 0(1/\sigma)$$

so that the general time-symmetric sourceless case is now reduced to the asymptotically Schwarzschildian time-symmetric sourceless case. In other words, we now have

$$g_{ij} = (1 + 2m/r)\,\delta_{ij} + 0(1/r^2), \qquad R[g] = 0$$

and want to show $m \geqq 0$.

For the proof, Schoen and Yau [11] assume the opposite, and show that in this case a 2-dimensional surface suitable for a 2 + 1 decomposition can be constructed; and with its help they derive a contradiction. I outline the main ideas of the argument. Let x, y, z be asymptotically Cartesian coordinates, and construct the minimal surface spanned by the circle $x^2 + y^2 = R^2$ in the xy plane. In this case of low dimension it is known that such a surface of strictly minimum area exists and is smooth. Furthermore, any maximum or minimum, z_0, of z must occur in the interior region. This is so because asymptotically the minimal surface condition becomes (with ${}^2\Delta$ the 2-dimensional Laplacian on the surface)

$${}^2\Delta z = -2mz_0/r + 0(1/r^3),$$

hence if $m < 0$, no maximum can occur in the asymptotic region at positive z_0, and no minimum at negative z_0. In fact, the surface must lie between two surfaces $\pm z = \text{const}$, where the constant is no larger than the radius beyond which the above asymptotic expression for ${}^2\Delta z$ dominates over the $0(1/r^3)$ terms. Therefore the radius R of the spanning circle can be expanded to infinity, and the surface will remain regular. Call the limit surface Σ. (In the case $m > 0$ the minimal surface would exist for any finite R, but it would "run away" to infinity in z as $R \to \infty$).

Now vary Σ by a one-parameter (s) family of surfaces. Since Σ is minimal its area A must satisfy

$$\mathrm{d}A/\mathrm{d}s = 0, \qquad \mathrm{d}^2A/\mathrm{d}s^2 \geqq 0.$$

This second variation can also be computed generally in terms of curvature quantities (superscript 2 denotes 2-dimensional quantities intrinsic to Σ) and the normal n^i to Σ,

$$\mathrm{d}^2A/\mathrm{d}s^2 = -\int (R_{ij}n^in^j - {}^2K_{ij}{}^2K^{ij} - {}^2K^2)\,\mathrm{d}A$$

Into this substitute the 2 + 1 decomposition (Gauss equation) of the constraint $R[g] = 0$,

$${}^2R = -{}^2K^{ij}\,{}^2K_{ij} + {}^2K^2 - 2R_{ij}n^in^j,$$

and use the Gauss-Bonnet theorem on Σ to find

$$\mathrm{d}^2A/\mathrm{d}s^2 = -1/2 \int {}^2K_{ij}{}^2K^{ij}\,\mathrm{d}A < 0,$$

which contradicts the minimal area property of Σ.

Of course the rigorous proof is considerably more complicated; for example, the Gauss-Bonnet theorem for Σ, which was here simply taken to be $\int {}^2R\,\mathrm{d}A = 0$, requires considerable justification.

5. Proof of Jang

We follow the previous argument to the end of section 3 and consider another type of 2 + 1 decomposition which turned out to be successful in proving the positivity of total energy of sourceless, time-symmetric gravitational waves. In addition, this method shows up the connection between positive energy, horizons, and cosmic censorship. Hawking [12] originated this method in his attempt to prove the positivity of Bondi energy. Geroch used similar methods in his argument for the positivity of ADM energy of time-symmetric gravitational waves [13]. Geroch's argument failed to be a proof because he had to assume the existence of a solution to a certain partial differential equation. Recently, Jang [14] found a modification of Geroch's argument in which one does not need such an assumption.

5.1. Case of Spherical Symmetry

The essential structure of this 2 + 1 decomposition can be best illustrated in spherically symmetric, asymptotically flat space-times. Introduce the usual coordinate system (t, r, θ, Φ) so that the metric takes the form

$$\mathrm{d}s^2 = -\mathrm{e}^{2\psi}(1 - 2M/r)\,\mathrm{d}t^2 + (1 - 2M/r)^{-1}\,\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\Phi^2) \tag{4}$$

where M and ψ are functions of r and t only, and $r > 2M$. Using energy dominance, Eq. (3), and the constraints, Eq. (2), one can then show

$$(\hat{t}^i + \hat{r}^i)\,\partial_i M \geqq 0 \qquad \text{and} \qquad (-\hat{t}^i + \hat{r}^i)\,\partial_i M \geqq 0$$

where $\hat{t}^i$ and $\hat{r}^i$ are unit vectors along t and r direction, respectively. That is, the M associated with each 2-sphere (r = const.) increases monotonically as one moves outward from one 2-sphere to the next along spacelike or null directions. Hawking noticed that this function m can be expressed more geometrically as the integral over each 2-sphere,

$$M = A^{1/2}/32\pi^{3/2} \int ({}^2R - 4\mu\varrho)\,\mathrm{d}A \tag{5}$$

Here A is the area of the integration surface, ϱ is the convergence of the outer null normal l^i, and $(-\mu)$ is the convergence of the inner null normal n^i, in the notation of Newman and Penrose. Quantities referring to the 2-sphere, such as its scalar curvature ${}^2R$, are denoted by preceding superscripts.

In the asymptotic region, both ϱ and μ are negative. Suppose that the gravitational field in the interior region is so strong that, as we move inward from the asymptotic region, ϱ or μ becomes zero on a certain 2-sphere H, of area A. From Eq. (5) and the Gauss-Bonnet theorem, we find that, on H, M assumes the value $(A/16\pi)^{1/2} = r/2$. (Hence, the line element of equation (4) becomes singular on H.) Since M increases as we move outward in any spatial direction, the value of M evaluated at spatial infinity, i.e. the ADM energy m, is not only positive but also greater than[1]) $(A/16\pi)^{1/2}$.

We need only a minor modification to deal with spherically symmetric space-times with no such surface H. In this case, the coordinate system of Eq. (4) can cover the whole space-time, and the regularity of the metric at the center of the spherical symmetry implies that M vanishes there. Hence, the monotonic property of M still allows us to conclude that the value of M evaluated at null infinity or at spatial infinity (i.e. the Bondi energy or the ADM energy, respectively) is non-negative.

5.2. General Case

For any time-symmetric slice Geroch [13] gave a positive energy argument which is a direct generalization of the above spherically symmetric one. Let H be the outermost minimal 2-sphere on a time-symmetric slice S. Such an H is called an apparent horizon. Geroch introduces a family of topological 2-spheres, parameterized by $r \in (r_0, \infty)$, in such a way that $r = r_0$ on H, and as r increases to infinity the surfaces approach metric 2-spheres in the asymptotic region. Next, on each 2-sphere, set

$$M = (A^{1/2}/32\pi^{3/2}) \int ({}^2R - {}^2K^2/2)\, \mathrm{d}A\,.$$

Note that this expression for M can be obtained from that of equation (5) using the fact that S is a time symmetric slice.

Geroch then assumes that a one-parameter family of 2-surfaces can be chosen such that the following equation holds on each surface,

$${}^2K = 2(\partial_i r \partial^i r)^{1/2}/r\,.$$

It is then a straightforward computation to show that $\mathrm{d}M/\mathrm{d}r \geqq 0$ for $r > r_0$. Since $M = (A/16\pi)^{1/2}$ on H, M evaluated in the asymptotic region, i.e., the ADM energy m, is greater than $(A/16\pi)^{1/2}$. As in the spherically symmetric case, we need only minor modifications to deal with time-symmetric slices without an apparent horizon. In this case the 2-surfaces should be chosen so that they reduce to a point as r approaches zero. Then, noting that $M \to 0$ as $r \to 0$ and the montonicity of M, one can still conclude the positivity of m.

Even though intuitive arguments suggest that Geroch's 2-spheres generally exist, this has not yet been proved.[2]) Recently Jang [14] modified Geroch's argument to prove the positivity of m for maximal slices. However, for slices with an apparent horizon, Jang's method does not prove $m \geqq (A/16\pi)^{1/2}$ but only the positivity of m. We briefly sketch Jang's proof.

---

[1]) The relation $m > (A/16\pi)^{1/2}$ also follows from the cosmic censorship hypothesis and other commonly held beliefs concerning gravitational collapse. Penrose [15] proposed to test the cosmic censorship hypothesis by examining the validity of such relationships. The above shows that spherically symmetric space-times do not violate this consequence of cosmic censorship; a larger class of non-violating space-times was given by Jang and Wald [16].

[2]) Since this argument seems to be the most promising in establishing the above mentioned relationship between m and the area of the apparant horizon, it would be of interest to settle this existence question in one way or the other. (In this regard, we note that it is not at all clear how one can modify Schoen and Yau's method to prove $m \geqq (A/16\pi)^{1/2}$ even for the spherically symmetric case.)

Let f be the solution of the Laplace equation with the boundary conditions that $f = f_0$ at H (where f_0 is a positive constant) and $f = 0$ at infinity. Let the flux integral of $\partial_i f$ over a 2-surface enclosing H be $(-4\pi e)$, and define r by e/f. On each $r =$ constant surface, set

$$W = (1/16\pi) \int \left[r({}^2R - {}^2K^2/2) + (r/2)\left({}^2K - 2(\partial_i r \partial^i r)^{1/2}/r\right)^2\right] \mathrm{d}A$$

If the $r =$ constant surface is a submanifold, $\mathrm{d}W/\mathrm{d}r$ evaluated on that surface is non-negative. As before, we note that W is positive on H and that W evaluated at infinity is also the ADM energy m; hence we can conclude the positivity of m. However, there are more technical details involved, because the level surfaces of the harmonic function f need not be submanifolds. Jang handles this difficulty by using a non-degenerate function f' which is arbitrarily close to f in a suitable sense.

Showing the positivity of energy is perhaps the first step in understanding the relationship between the interior and the asymptotic structure of a space-time. I believe that the various methods developed in the positive energy program will find many interesting applications in the study of isolated systems in general relativity.

References

[1] Dieter R. Brill and Pong Soo Jang, "The Positive Mass Conjecture" in A. Held (Ed.), General Relativity and Gravitation 100 years after the birth of Einstein, Plenum Publishing Co, New York 1980.
[2] A. Ashtekar in A. Held (Ed.), loc. cit.
[3] See, for example, Y. Choquet-Bruhat and J. York, "The Cauchy Problem" in A. Held (Ed.), loc. cit.
[4] D. Christodoulou and N. O'Murchadha, GR 9 contributed paper
[5] Y. Choquet-Bruhat and J. Marsden, C. R. Acad. Sci. Paris **282**, 609 (1976).
[6] N. O'Murchadha and J. York, Phys. Rev. **D 10**, 2345 (1974).
[7] R. Schoen and S.-T. Yau, Phys. Rev. Letters **43**, 1457 (1979).
[8] P.-S. Jang, J. Math. Phys. **19**, 1152 (1978).
[9] D. Brill, Ann. of Phys. (N.Y.) **7**, 466 (1959).
[10] J. York, "Has the Positive Energy Conjecture been proved?", Preprint (1979).
[11] R. Schoen and S.-T. Yau, Commun. Math. Phys. **65**, 45 (1979); Phys. Rev. Letters **43**, 1457 (1979); Preprint (1980).
[12] S. Hawking, Gravity Prize essay (1970).
[13] R. Geroch, Ann. N.Y. Acad. Sci. **224**, 108 (1973).
[14] P.-S. Jang, Commun. Math. Phys. **69**, 257 (1979).
[15] R. Penrose, Ann. N.Y. Acad. Sci. **224**, 125 (1973).
[16] P.-S. Jang and R. Wald, J. Math. Phys. **18**, 41 (1977).

Address: Prof. Dr. Dieter R. Brill
University of Maryland
College Park, Md. 20782
USA

Prof. Dr. P.-S. Jang
Syracuse University,
Syracuse, N.Y. 13210
USA

Present Status of Gravitational-Wave Experiments[1])

V. B. Braginsky (Moscow) and K. S. Thorne (Pasadena)

We welcome you to our joint Soviet-American song and dance about the beauty of recent experimental work on gravitational radiation.

Our show consists of five parts: a review of current astrophysical predictions of the strengths of gravity waves bathing the Earth (§ 1); a description of how the challenge to detect gravity waves is creating a new chapter in the field of quantum electronics (§ 2); and descriptions of the current status of three types of gravity-wave detection systems: Weber-type bars (§ 3), laser interferometers (§ 4), and spacecraft tracking (§ 5).

1. Predicted strengths of gravity waves

Astrophysicists and relativity theorists have worked hard, during the past five years, to make realistic estimates of the strengths and frequencies of the gravity waves bathing the Earth. This research was highlighted by a two-week workshop on "Sources of Gravitational Radiation" in Seattle, Washington, USA, July 4—August 28, 1978. [1] Many sources were studied: Possible sources of broad-band bursts included the collapsing and bouncing cores of supernovae in our galaxy and other galaxies; neutrinos pouring out of a supernova; corequakes in neutron stars; the births of black holes ranging from 3 solar masses to 10^9 solar masses; collisions between black holes and between black holes and neutron stars, which may occur in globular clusters and in galactic nuclei and quasars; and the final inspiral, coalescence, and destruction of compact binaries such as the binary pulsar. Possible sources of periodic waves included binary star systems; rotating, deformed neutron stars; rotating, deformed white dwarfs; and pulsations of white dwarfs that may follow nova outbursts. Possible sources of a stochastic background included the big-bang singularity, inhomogeneities in the very early universe, and the deaths-to-form-black-holes of "Population III stars" (stars that were born before galaxies formed).

From the many studies of all these sources and more, one of us [2] has drawn the rather subjective conclusions shown in fig. 1. There we plot vertically the dimensionless amplitude h of the gravitational waves (magnitude of the transverse-

[1]) Supported in part by the Ministry of Higher Education, USSR, and by the National Science Foundation, USA.

traceless part of the metric perturbation associated with the waves), and horizontally the frequency of the waves. Broad-band bursts, arriving at Earth once per month, could have amplitudes as large as the topmost line in fig. 1 without violating any conventional "cherished beliefs" about the nature of gravity or the astrophysical structure of our Universe. However, currently fashionable models of the Universe

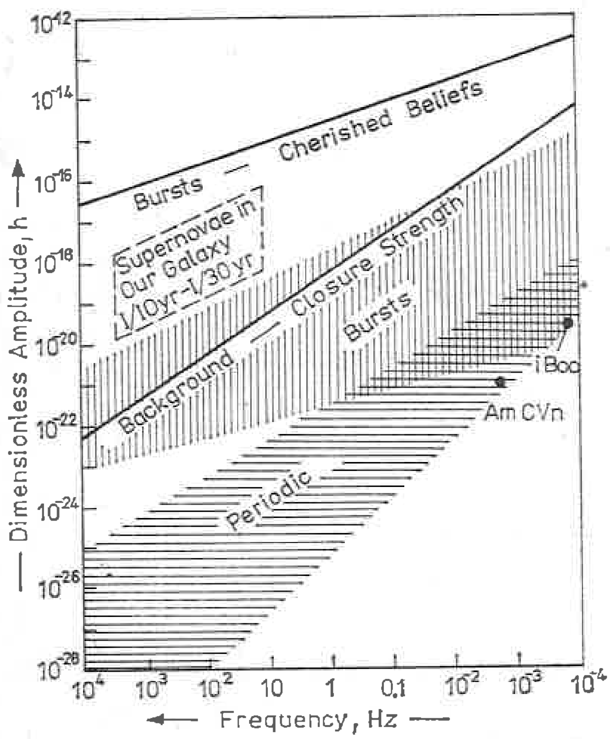


Fig. 1. Estimates of the strengths of the gravitational waves that bathe the Earth. See text for explanation of the lines and hatched regions (from reference [2]).

suggest that the strongest once-per-month bursts might lie far below the "cherished belief" line — somewhere in the vertically hatched region of fig. 1. The first burst to be discovered may well lie above the once-per-month hatched region — for example, it might be from a supernova in our Galaxy with strength somewhere in the dashed region of fig. 1.

Similarly, fashionable models for the Universe suggest that the strongest periodic sources might lie in the horizontally hatched region of fig. 1. And there could exist a stochastic background as large as the solid line (energy density enough to close the Universe) — though some strong but not certain astrophysical arguments suggest the background will be weaker than this by several orders of magnitude or more. [3]

Figure 1 shows an enormous range of possible wave strengths, corresponding to our enormous ignorance about the existence of and astrophysical behaviors of the sources. Our ignorance results from the fact that the astrophysical information carried by gravity waves is nearly orthogonal to the information we now receive from radio, infrared, optical, ultraviolet, X-ray, and gamma-ray observations. Because of this orthogonality, gravity waves have the potential to become a powerful tool for astronomy — a clear window onto parts of the Universe that we now perceive hardly at all. We have no guarantee that the gravity-wave window will be opened successfully during the next decade, or even before the year 2000. But prospects for success in the 1980's are reasonably good.

We can summarize by saying that the program to detect gravity waves is one of moderately high risk and enormously high potential payoff.

2. A new chapter in quantum electronics

Figure 1 suggests a long-term goal of $h \sim 10^{-21}$ for ground-based detectors, which operate at frequencies $f \sim 100$ Hz to 10^4 Hz. It is not ridiculous to hope to be nearing this goal by the end of the 1980's. To reach this goal will require measuring displacements of the ends of a one-meter bar detector with precision

$$\Delta x_{\mathrm{GOAL}} \simeq h \cdot (1 \text{ meter}) \simeq 1 \times 10^{-19} \text{ cm}. \quad (1)$$

For comparison, if one regards the fundamental mode of oscillation of such a bar as a quantum mechanical oscillator of mass $m = 1$ ton and frequency $\omega/2\pi = 10^3$ Hz, the half-width of its ground-state wave function is

$$\Delta x_{\mathrm{GS}} = (\hbar/2m\omega)^{1/2} = 3 \times 10^{-19} \text{ cm} = 3\Delta x_{\mathrm{GOAL}}. \quad (2)$$

Thus, one-ton detectors of the late 1980's must be regarded as quantum mechanical oscillators. By contrast, the gravity waves one seeks to detect are highly classical — they typically have graviton occupation numbers, averaged over the beam-width of the antenna, of order 10^{37}.

Classically one describes the oscillations of the bar's fundamental mode by the displacement x of the bar's end from equilibrium and by the corresponding canonical momentum p. As time passes x and p oscillate sinusoidally

$$\begin{aligned} x &= X_1 \cos \omega t + X_2 \sin \omega t, \\ p/m\omega &= -X_1 \sin \omega t + X_2 \cos \omega t, \end{aligned} \quad (3)$$

where X_1 and X_2 are constant amplitudes for the two different phases of oscillation. Quantum mechanically x and p are Hermitean operators, and so also are X_1 and X_2. The commutation relation $[x, p] = i\hbar$ implies $[X_1, X_2] = i\hbar/m\omega$, which in turn implies the Heisenberg uncertainty relation [4, 5]

$$\Delta X_1 \Delta X_2 \geqq \hbar/2m\omega \equiv (\mathrm{SQL})^2. \quad (4)$$

Here $\mathrm{SQL} \equiv \Delta x_{\mathrm{GS}} = (\hbar/2m\omega)^{1/2} \simeq 3 \times 10^{-19}$ cm is the so-called "standard quantum limit" for a gravity-wave detector. In real experiments, with real mechanical/electrical transducers followed by electronic amplifiers, $\hbar$ is replaced by $2kT_a/\Omega$ where k is Boltzman's constant, T_a is the amplifier noise temperature, and Ω is the frequency at which the amplifier operates. [6, 7] Then the standard quantum limit gets replaced by a standard amplifier limit

$$\Delta X_1 \, \Delta X_2 \geqq \frac{kT_a/\Omega}{m\omega} \equiv (\mathrm{SAL})^2 \geqq (\mathrm{SQL})^2. \tag{5}$$

In all past and present bar-type detectors the electronics were designed to measure X_1 and X_2 with equal precision (neither phase preferred over the other; circular error box in phase plane of fig. 2). Such measurements can never measure gravity-induced amplitude changes more accurately than the SAL — which is inadequate for the long-range goals of gravity-wave astronomy.

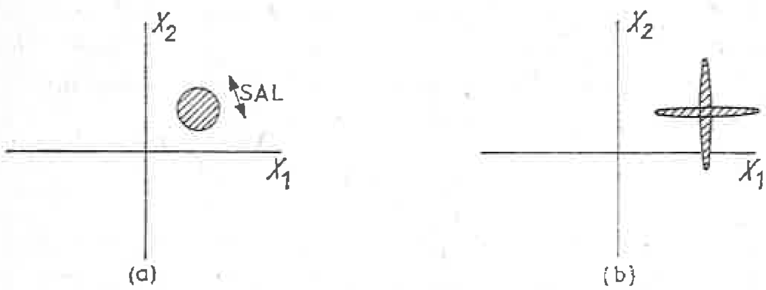


Fig. 2. Uncertainty error boxes for the amplitudes (X_1, X_2) of a bar-type detector of gravitational waves. The error box for past and present measuring systems (drawing (a)) is round and has a diameter no smaller than the standard amplifier limit SAL. Quantum nondemolition measuring systems, planned for the mid and late 1980's, produce long, thin error boxes (drawing (b)).

This obstacle was recognized early in the 1970's as a result of detailed analyses of the limiting performances of various measuring schemes. [8, 9] The above elementary viewpoint (Eqs. 3 to 5) brought with itself, in late 1977, an obvious solution to the obstacle. [5, 6] Construct two detectors. On one detector measure X_1 with high accuracy ($\Delta X_1 \ll \mathrm{SAL}$) — and in the process perturb X_2 so badly that one can gain little information about it ($\Delta X_2 \gg \mathrm{SAL}$). On the second detector measure X_2 with high accuracy ($\Delta X_2 \ll \mathrm{SAL}$), thereby perturbing X_1 ($\Delta X_1 \gg \mathrm{SAL}$). On a common plot of the error boxes of the two oscillators (fig. 2) their tiny intersection point will move, under the influence of a gravity wave, in precisely the same manner as the system point of a classical oscillator. As a result, one can measure the full and detailed effects of the classical gravity wave with the desired precision $\Delta x \simeq \frac{1}{3}$ SQL — and in principle one can measure them with arbitrary precision, despite the quantum mechanical nature of one's detectors. The proposal to make such measurements, and specific (rather simple) designs of apparatus for doing so, are opening up a new chapter of quantum electronics called "Quantum Nondemolition Measurements" (QND). For reviews see references [7, 10, and 11].

Just as QND will likely be a key element in bar detectors of the mid and late 1980's, so a second new concept — "squeezing the vacuum" [12] will likely be a key element in laser-interferometer detectors. Recall that in such a detector one uses laser interferometry to measure the armlength difference $L_1 - L_2$ in a system consisting of

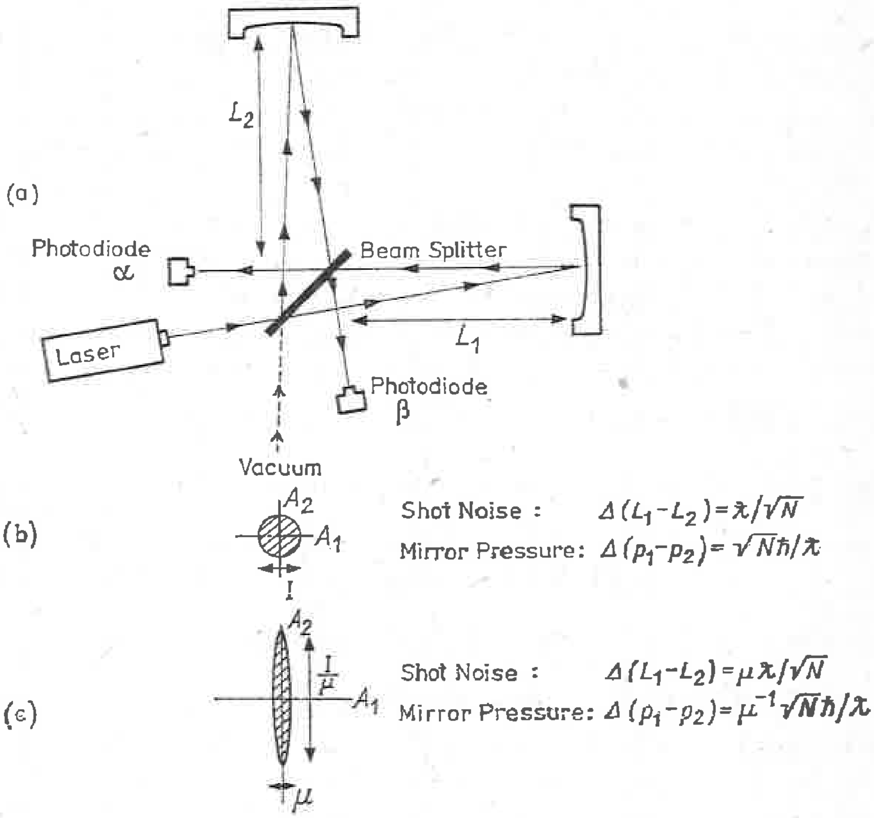


Fig. 3. (a) Idealized version of a laser-interferometer gravity-wave detector.
(b) The uncertainty error box for a normal mode of the electromagnetic field in its ground state (vacuum fluctuations). The superposition of these vacuum fluctuations on the laser light produces the indicated noises in the interferometer.
(c) The uncertainty error box for a squeezed state of the electromagnetic field, and the resulting superposition noises in the interferometer.

three masses suspended as pendula from an overhead support (fig. 3a). Such measurements have two ultimate sources of noise:

(i) *photon shot noise*, which for the simple "one-bounce" interferometer of fig. 3a gives the limit

$$\Delta(L_1 - L_2) \simeq \lambda\!\!\!^-/\sqrt{N} \tag{6a}$$

with $\lambda\!\!\!^- =$ (wavelength of light)/2π, $N =$ (number of photons collected by photodiodes during averaging time τ); and (ii) *photon pressure fluctuations* on the mirrors, which produce differential momentum changes in the two arms [13]

$$\Delta(p_1 - p_2) \simeq \sqrt{N}\,\hbar/\lambda\!\!\!^{-} \simeq \hbar/\Delta(L_1 - L_2) \tag{6b}$$

and consequent length changes

$$\Delta(L_1 - L_2) \simeq (\tau/m)\,\Delta(p_1 - p_2) \tag{6c}$$

during the averaging time τ. With currently available laser powers of $\lesssim 10$ watts the shot noise (6a) is much larger than the pressure-fluctuation noise (6b, c). One can reduce the shot noise (at the price of increasing the pressure fluctuations) by either of two methods: invent a more powerful laser (make N larger), or "squeeze the vacuum".

Squeezing the vacuum (an idea invented two months ago by Carlton Caves [12]) can be understood as follows: Both the shot noise and the pressure fluctuations can be regarded as due to a superposition of two electromagnetic fields in the interferometer: the laser field which enters from the left part of fig. 3a, and the vacuum field which enters from the bottom part. The vacuum field has equal amounts of noise ($\frac{1}{4}$ quantum) in each of its two phases A_1 and A_2. The noise in phase A_1 superimposes on the laser light (N quanta) to give a random difference $\Delta(N_\alpha - N_\beta) \simeq \sqrt{\frac{1}{4} \cdot N} \simeq \sqrt{N}$ in the number of quanta collected by the two photodiodes, resulting in the shot noise limit (6a). The noise in phase A_2 superimposes on the laser light to give a random difference $\Delta(N_1 - N_2) \simeq \sqrt{\frac{1}{4} \cdot N} \simeq \sqrt{N}$ in the number of quanta bouncing off the two mirrors, resulting in the pressure fluctuation limit (6b, c). By sending the vacuum electromagnetic field through a nonlinear optical device called a "degenerate parametric amplifier", one can squeeze its uncertainty error box by a factor μ (fig. 3c) before it enters the interferometer, thereby reducing the mean number of quanta in A_1 to $\mu^2/4$ and increasing the number in A_2 to $\mu^{-2}/4$, and thence reducing the shot noise by a factor μ while increasing the pressure fluctuations by a factor μ^{-1}. This technique may well be used in the mid and late 1980's, unless lasers of much increased power are invented.

The ideas of quantum nondemolition and of squeezing the vacuum are likely to find application not only in gravitational wave detection, where they were invented, but also in a variety of other areas of modern technology. Even if gravitational waves are never discovered, the effort may well be justified by technology spin-offs such as these.

3. Weber-type bar detectors

The use of resonant bars as gravity-wave detectors was pioneered by Joseph Weber at the University of Maryland during the 1960's; and detectors of similar design were constructed in Moscow, Glasgow, Frascati, Munich, Bell Labs, Rochester, IBM, Tokyo, Bristol, Reading, Stanford, LSU, Rome, Meudon, and Regina during the 1970's. By 1975 the most sensitive of these room-temperature, aluminum Weber-type bars had achieved rms noise levels of $h \simeq 1 \times 10^{-16}$ for frequencies $f \simeq 10^3$ Hz. For a review see, e.g., [14].

The GR 8 meeting in Waterloo (July 1977) occurred just two years after the completion of these "first-generation" detectors, and at a time when the design and early development work for a "second generation" was underway. Since GR 8 good progress has been made on the second generation, in which the bars are made of new types of materials and are cooled to 2 °K or less, and their oscillations are measured by totally new types of mechanical/electrical transducers. One such detector (at Stanford) is now fully in operation, though with an rms sensitivity $h \simeq 3 \times 10^{-18}$ that is well below its ultimate goal. Others elsewhere can be expected to start operating within the next year or two.

Table 1 summarizes the nature and current status of the bar-detector research in various laboratories. Column I of Table 1 shows the location of each laboratory and lists references describing that laboratory's recent work. Column II lists the mass of the bar to be used in the final version of the detector. (Some laboratories prefer to use lower-mass, prototype bars in their present developmental work; the prototype masses are not listed.) Column II also lists the temperature of the bars in current detectors. The statement "< 2 °K" means that this laboratory possesses refrigeration equipment capable of cooling the bars to temperatures below 2 °K. The bars now being used or now being negotiated for purchase are made of various materials shown in column III. Here "Al" means the same type of aluminum as was used in first-generation bars; "5056" means a special alloy of aluminum with large magnesium content, which the Tokyo group [29] has discovered to have a rather high mechanical Q; "Si" means silicon; "Nb" means niobium; and "Al_2O_3" means sapphire. Column IV shows the mechanical quality factor Q of current prototype bars when cooled to the operating temperature of ~ 2 °K. Column V describes the transducer system and associated electronics being developed. Column VI shows the displacement sensitivity actually achieved as of July 1980, with a prototype transducer with a 1 Hz bandwidth (1 second averaging time) — except that numbers shown in parentheses and marked * are not sensitivities actually achieved but rather estimates of sensitivities that would result if the present transducer were coupled to the best existing DC SQUID amplifier (or, in the LSU case, to the existing SUPRAMP). Column VII lists the sensitivity (RMS noise level) that would result today if a transducer which has actually operated were coupled to the planned bar. In one case, that of Stanford (last row), the coupling of transducer to bar has actually been carried out and the resulting detection system is fully operational.

Note that Stanford's gravity-wave sensitivity (column VII) is roughly a factor 30 better than the sensitivities of the best first-generation bars. The ultimate goal of second-generation work is improvement by an additional factor of 100, to get near the standard amplifier limit for a several-ton bar: $h \simeq 3 \times 10^{-20}$.

Comparison with fig. 1 shows that current detectors are more sensitive than the cherished-belief upper limit: Nothing in our current views of physics or the Universe forbids the Stanford detector to see many bursts per month. Moreover, current sensi-

Table. 1 *Present instrumentation and status of second-generation Weber-type bar detectors in various laboratories (July 1980)*

I Location of lab. & refs.	II Mass & temperature	III Material of bar	IV Q of bar	V Mechanical/electrical transducer & electronics	VI Transducer sensitivity actually achieved for 1 sec. averaging time	VII Equivalent sensitivity in h (present state-of-the-art)
Maryland 15, 16, 17, 18, 19	3×10^6 g < 2°K	5056 Si	5×10^7	2-stage mechanical diaphragm + Inductive transducer + DC SQUID	(3×10^{-18} cm)*	
Perth 20, 21	1.6×10^6 g 2°K	Nb	6×10^7	Microwave-cavity capacitive readout system	1.4×10^{-15} cm	2×10^{-17}
Rome-CERN-Frascati 22, 23	5×10^6 g < 2°K	Al 5056	1×10^6	Inductive transducer + Microwave SQUID	1.5×10^{-15} cm	1×10^{-17}
Rochester 24	2×10^6 g 1°K	Al 5056	8×10^7	Inductive transducer + SQUID	(1×10^{-17} cm)*	
Louisiana State Univ. 25	5×10^6 g < 2°K	Al 5056	1.5×10^6	Microwave-cavity transducer + Supramp	(2×10^{-16} cm)*	
Moscow 26, 27	1.5×10^4 g 2°K	Al_2O_3	5×10^9	Microwave-cavity capacitive readout system + QND	3×10^{-16} cm	5×10^{-18}
Stanford (operating now) 28	5×10^6 g 4°K	Al	5×10^6 (1×10^5 loaded)	1-stage mechanical diaphragm + Inductive transducer + Microwave SQUID	6×10^{-16} cm (operating now)	3×10^{-18} (System noise temperature is 0.01°K)

tivities are entering the realm of strongly aspherical supernovae anywhere in our Galaxy.

In addition to the kilohertz-band detectors described in Table 1, there is at the University of Tokyo [30] a project to construct a mechanically resonant antenna for detecting periodic gravity waves from the Crab pulsar. The main goal is to reach a sensitivity of $h \lesssim 10^{-25}$ — corresponding to the most optimistic current estimates of the Crab-pulsar wave strength. (The "best guess" current estimate is $h \simeq 3 \times 10^{-27}$.) This project involves deep cooling of a massive antenna, and a long averaging of the signal so as to reach a sensitivity to periodic sources far better than one can achieve for burst sources.

4. Laser-interferometer detectors

Laser interferometers offer a design approach to Earth-based gravity-wave detection (fig. 3a) which is very different from the bar design. It is likely that both designs will be used in the long run — each proving superior to the other for gravity waves of a specific temporal character. Laser systems have the advantage of being intrinsically broad band (i.e., a laser system can operate over more than a decade of frequencies all at once, thereby studying the detailed time structure $h(t)$ of the gravity wave). All current bar detectors are narrow band; and, although they can be made broad band in principle, to do so will be difficult. Laser systems have the disadvantage of being technologically more complex and more expensive than bars.

The first prototype laser system was constructed and operated in 1972 at Hughes Research Laboratories, [31] with an rms sensitivity of $h \sim 10^{-14}$ for $f \simeq 1$ to 10 kHz. Simultaneous with this demonstration experiment, a more sophisticated laser system was under construction at MIT. [32] In 1975 vigorous laser-system efforts were initiated at Munich and Glasgow and in 1979 at Caltech.

The strategy of all these efforts is based on the hope that, once sophisticated prototypes of modest length have been operated successfully, the sensitivity $h = \Delta(L_1 - L_2)/L_1$ to gravity waves can be improved fairly rapidly by scaling up the length $L_1 \simeq L_2$ of the arms, without making major changes in the instrumentation by which the length difference $L_1 - L_2$ is monitored. For this reason, all effort thus far has focussed on developing the monitoring instrumentation on prototype detectors of modest arm length: 1 meter at MIT, 3 meters in Munich, 1 meter and 10 meters in Glasgow. Much attention is paid to identifying all noise sources, understanding them thoroughly, and devising ways to remove them which will work not just on the current prototypes, but also on much larger future detectors.

The noise sources are so many and so varied, and the possible ways of removing them are so numerous, that there is a "near infinity" of different possible detector designs. Each group is pursuing a design rather different from the others (for example, MIT and Munich are constructing Michelson interferometers in which a pencil beam bounces many times off each mirror, making many discrete and independent spots on the mirror; Glasgow is constructing a Fabry-Perrot interferometer in which each arm is excited as a giant, optically resonant cavity with just one spot on each mirror, but with the photons still bouncing back and forth many times). Despite the numerous differences in design, there is a great deal of overlap of technology and technique from one laboratory to the next. Whenever experimenters from different laboratories meet to discuss their progress, they learn unexpected, crucial things from each other. And, of course, this is just as true of bar experimenters as it is of interferometer experimenters.

The Munich prototype [33, 34] has recently operated successfully at the photon shot-noise limit $\Delta(L_1 - L_2) \simeq \lambdabar/\left(B\sqrt{N}\right)$ [where $\lambdabar$ = (wavelength of light)/2π $\simeq 1 \times 10^{-5}$ cm; B = (number of bounces of light in interferometer) = 120; N = (number of photons collected from 25 milliwatt beam during photon averaging time of $\tau = 3 \times 10^{-4}$ sec.) $\simeq 3 \times 10^{13}$], for gravity waves of frequency $f \simeq 1/2\tau \simeq 2$ kHz. The corresponding gravity-wave sensitivity, $h \simeq 1 \times 10^{-16}$, is comparable to that of the first-generation bar detectors. The MIT and Glasgow [35] detectors are not far from similar operation; and we can hope for rather rapid improvements over the next few years.

The ultimate goal during the early and mid 1980's is to construct detectors with arm lengths $L \sim 10^5$ to 10^6 cm, with mirrors of high enough reflectivity to permit $B \sim 500$ bounces, and with the most powerful continuous lasers now available (several watts). Such detectors should be comparable in sensitivity to the best anticipated second-generation bar detectors $h \simeq 3 \times 10^{-20}$. Further improvements in the longer run — using larger laser powers, higher mirror reflectivities, and Caves' "squeezing of the vacuum" — might give sensitivities approaching the "standard quantum limit" for a laser-interferometer detector:

$$\text{SQL}: h \simeq \frac{\Delta(L_1 - L_2)}{L} \simeq \frac{(\hbar\tau/m)^{1/2}}{L} \lesssim 1 \times 10^{-22} \tag{7}$$

[limit obtained by adjusting N in Eqs. (6) so that photon shot noise (6a) and photon pressure noise (6a) are equal]. Here τ is the photon averaging time, which must not exceed $1/2f$ with $f \sim 100$ Hz to 10^3 Hz the gravity-wave frequency; $m \simeq 10^4$ g is the mirror mass; and $L \simeq 10^5$ cm to 10^6 cm is the arm length. This SQL, $h \lesssim 1 \times 10^{-22}$, is far smaller than the corresponding SQL for bar detectors, $h \sim 10^{-19}$, because the detector length L is so much bigger.

Of course, starry-eyed dreams induced by the smallness of this limit must be tempered by the realities of current sensitivities ($h \sim 10^{-16}$) and of the enormous technological hurdles between here and there.

5. Doppler-tracking of spacecraft

Seismic noise is filtered away from bar and laser-interferometer detectors without much difficulty. However, at low gravity-wave frequencies ($f \lesssim 1$ Hz) such filtering is exceedingly difficult if not impossible; and at these frequencies, fluctuating gravity gradients due to people, animals, automobiles, trucks, airplanes, ... also create prohibitively large noise. To avoid these noises one must use detectors in space.

Doppler tracking of spacecraft is the most sensitive space technique now available. Although its use for gravity-wave detection was proposed in the mid-1960's, a serious effort at using it was not initiated until after GR8. In October 1977 and January 1978 the research group at JPL collected high-precision Doppler data from the Viking spacecraft in orbit around Mars, and used those data to identify and characterize the dominant noise sources in the present gravity-wave detection system. [36] By far the dominant noise was due to fluctuations in the interplanetary plasma, which cause fluctuations of dispersion in the "S-band" ($\sim 2 \times 10^9$ Hz) radio tracking signal and thereby cause spurious Doppler-shift fluctuations of magnitude $\Delta\nu/\nu \simeq 3 \times 10^{-13}$ to 3×10^{-14} on timescales 10^3 to 10^4 seconds.

The first attempt to actually detect gravity waves with the current Doppler tracking system was made this year at JPL using the Voyager spacecraft. [37, 38] The Voyager data yielded an upper limit of $h < 3 \times 10^{-14}$ for bursts of characteristic timescale 500 seconds, during a total tracking time of one day. They also yielded an upper limit of $[f \times S_h(f)]^{1/2} < 3 \times 10^{-14}$ on the spectral density $S_h(f)$ of any stochastic background of gravity waves at frequencies 10^{-4} Hz $< f < 10^{-2}$ Hz. For comparison, from seismometer studies of the Earth's vibrations, Joseph Weber [39] in 1967 was able to place the limit $[f \times S_h(f)]^{1/2} < 1 \times 10^{-14}$ at the Earth's quadrupole vibration frequency $f = 3.1 \times 10^{-4}$ Hz.

NASA now possesses the capability to track spacecrafts with "X-band" ($\sim 8 \times 10^9$ Hz) radio signals and to thereby reduce the fluctuations of plasma dispersion by more than a factor 10. The Viking studies [36] suggest that no other noise sources will prevent gravity-wave sensitivities from thereby improving by a factor 10; and even if those studies are wrong and other noise sources (e.g., dispersion in the Earth's troposphere) are unexpectedly large, the improved system with 100 days of tracking should at least be able to improve the sensitivity to a stochastic background by a factor 10. [40] Although the "politics" of doing so are not yet fully settled, it seems likely that NASA will fly the improved Doppler system on its Solar Polar spacecraft in the mid 1980's; and the gravity community is pushing hard for the improved system to be flown then also on a Galileo spacecraft or on a European Solar Polar spacecraft, so that coincidence experiments can be performed. It is not impossible that this system will detect gravity-wave bursts from supermassive black holes in galactic nuclei at the Hubble distance, and that the first detections will be reported at the GR11 conference.

In the long run, perhaps in the late 1980's, NASA might fly a yet-more-sophisticated Doppler system, with sensitivity $h \lesssim 10^{-16}$. This system, now being designed at the Smithsonian Astrophysical Observatory by R. Vessot and colleagues, [41] would involve carrying a very stable microwave oscillator ("clock") on the spacecraft, and measuring Doppler shifts with reference to that clock as well as with reference to the present ground-based clock. This system would have four independent Doppler readouts: two one-way (Earth $\rightarrow$ spacecraft and spacecraft $\rightarrow$ Earth); two round-trip (Earth $\rightarrow$ spacecraft $\rightarrow$ Earth and spacecraft $\rightarrow$ Earth $\rightarrow$ spacecraft). The time correlations produced by a gravity wave in these four readouts would differ markedly from the time correlations produced by the various noise sources.

One key to further improvements in the Doppler system is the development of new clocks with higher frequency stability. Vessot [41] has recently made substantial improvements in the Hydrogen maser clock, pushing its level of frequency stability down to $\Delta\omega/\omega \simeq 6 \times 10^{-16}$ for averaging times of 3000 seconds; and the prospects for further improvement are excellent. Turneaure [42] at Stanford has achieved $\Delta\omega/\omega \simeq 1.6 \times 10^{-16}$ for averaging times of 200 seconds with a "superconducting cavity stabilized oscillator" clock. And as remarkable as these stabilities may be, they are still five orders of magnitude away from the ultimate quantum mechanical limit for the stability of microwave-frequency clocks. [43] Thus, it is not ridiculous to hope that Doppler-tracking sensitivities of $h \ll 10^{-16}$ will ultimately be achieved.

It may well be that in the 1990's experimenters will switch from radio tracking to optical (laser) tracking for space-borne gravity wave detectors, and will have several spacecrafts track each other, rather than tracking the spacecraft from Earth. Preliminary theoretical analyses [44, 45] of such optically-linked systems suggest that sensitivities $h < 10^{-21}$ might be achievable at all frequencies from $f \sim 30$ Hz (where Earth-based detectors might start cutting off) to $f \sim 10^{-4}$ Hz (where strong sources might start cutting off). Such a system could detect gravity waves from many binary stars, and would likely see a wide variety of other sources as well. However, before one can seriously propose instrument development for such a system, there must be far more detailed feasibility studies than have yet been carried out; and before such a system can fly there will have to be a long and challenging program of instrument development.

6. Conclusion

The worldwide program to detect gravitational waves has become a major effort, carried forward by the enthusiasm, energy, cleverness, and very hard work of more than one hundred scientists (Table 2). This program has produced technical ideas and inventions which are as intrinsically beautiful as exact solutions of Einstein's equations, and which may find extensive application elsewhere in science and technology. And the effort might even reach its goal, during the 1980's, of opening up the gravity-wave window onto the Universe.

Table 2. *A partial list of scientists now working on the effort to detect gravitational waves*

I. BAR DETECTORS

- *Maryland:* J. Weber, J. P. Richard, H.-J. Paik, W. Davies, K. Krack, G. Castellano, C. Cosmelli
- *Perth:* D. Blair, J. Bryant, M. Buckingham, J. Davidson, J. Ferrierino, C. Edwards, F. Van Rann, A. Mann, L. Mann, U. Veitch, R. James
- *Rome-CERN-Frascati:* E. Amaldi, G. Pizzella, I. Modena, G. Pallotino, F. Ricci, V. Ferrari, C. Cosmelli, S. Frasca, I. Bonifazi, F. Bordoni, F. Fuligni, V. Giovanardi
- *Rochester:* D. Douglass, W. Johnson, M. Karim, M. Bocko, R. Marsden, B. Muhlfelder, L. Narici, M. Cromar
- *Louisiana State University:* W. Hamilton, J. Kadlec, G. Wang, I. Campisi, G. Spetz, W. Oelfke
- *Moscow:* V. Braginsky, V. Mitrofanov, V. Panov, V. Rudenko, V. Popelnuk, E. Popov, A. Manukin
- *Stanford:* W. Fairbank, R. Giffard, R. Taber, P. Michelson, E. Mapols, C. Chun, M. McAshan
- *Bristol:* P. Aplin
- *Tokyo:* H. Hirakawa, K. Oide, K. Tsubono, M.-K. Fujimoto
- *Beijing:* Qin R., Hu R., Jiang N., Liu Y., Tan D., Tian J., Wang G., Zhang P., Zhao Z., Zheng L.
- *Guangzhou:* Chen J., Guan T., Huang Q., Lee Y., Qiu Z., Yang X., Yu P.

II. LASER INTERFEROMETER DETECTORS

- *MIT:* R. Weiss, P. Lindsay
- *Munich:* H. Billing, K. Maischberger, A. Rüdiger, R. Schilling, L. Schnupp, W. Winkler
- *Glasgow:* R. Drever, J. Hough, G. Ford, I. Kerr, A. Munley, J. Pugh, N. Robertson, H. Ward
- *Caltech:* R. Drever, S. Whitcomb, S.-A. Lee, R. Spero, M. Hereld, E. Brooks

III. SPACECRAFT TRACKING

- *JPL:* F. Estabrook, H. Wahlquist, R. Hellings
- *Upsala:* A. J. Anderson
- *Pavia:* B. Bertotti
- *Smithsonian:* R. Vessot
- *Marshall Spaceflight Center:* R. Decher, C. Lundquist
- *JILA:* P. Bender, J. Faller, J. Randall

References

[1] L. Smarr, Ed., Sources of Gravitational Radiation (Cambridge University Press, 1979).
[2] K. S. Thorne, Rev. Mod. Phys., **52**, 285 (1980).
[3] B. J. Carr, Astron. Astrophys., in press; and references reviewed therein.
[4] P. Kafka, in Proceedings of the International School of General Relativistic Effects in Physics and Astrophysics: Experiments and Theory (Third Course), Eds. R. Ruffini, J. Ehlers, and C. W. F. Everitt (Max Planck Institute für Physik und Astrophysik, Munich, 1977).
[5] K. S. Thorne, R. W. P. Drever, C. M. Caves, M. Zimmermann, V. D. Sandberg, Phys. Rev. Letters, **40**, 667 (1978).
[6] V. B. Braginsky, Yu. I. Vorontsov, and F. Ya. Khalili, Pis'ma Zhur. Eksp. Teor. Fiz., **27**, 296 (1978).
[7] V. B. Braginsky, Yu. I. Vorontsov, and K. S. Thorne, Science, **209**, 547 (1980).
[8] V. B. Braginsky, Physical Experiments with Test Bodies (Nauka, Moscow, 1970); English translation published as NASA-TT F762, National Technical Information Service, Springfield, VA.; see especially Eqs. (3.17) and (3.25).
[9] R. P. Giffard, Phys. Rev. **D 16**, 2478 (1976).
[10] K. S. Thorne, C. M. Caves, V. D. Sandberg, M. Zimmermann, and R. W. P. Drever, in: L. Smarr, Ed., Sources of Gravitational Radiation (Cambridge University Press, 1979), p. 49.
[11] C. M. Caves, K. S. Thorne, R. W. P. Drever, M. Zimmermann, and V. D. Sandberg, Rev. Mod. Phys., **52**, 341 (1980).
[12] C. M. Caves, Phys. Rev. **D**, submitted.
[13] C. M. Caves, Phys. Rev. Letters, **45**, 75 (1980).
[14] E. Amaldi and G. Pizzella, in: Relativity, Quanta, and Cosmology in the Development of the Scientific Thought of Albert Einstein, Vol. 1, Ed. F. De Finis (Johnson Reprint, New York), p. 9.
[15] J. P. Richard, in: Proceedings of the Einstein Centenary Summer School, Perth, Australia (1979).
[16] J. P. Richard and C. Cosmelli, GR9 Abstracts (ref. 46), p. 417 (1980).
[17] H.-J. Paik, GR9 Abstracts (ref. 46), p. 395 (1980).
[18] H.-J. Paik, Nuovo Cimento, **55B**, 15 (1980).
[19] J. Weber, University of Maryland Report 81-002 (1980).
[20] R. James, A. G. Mann, D. G. Blair, J. A. Davidson, and L. D. Mann, GR9 Abstracts (ref. 46), Vol. 2, p. 392.
[21] D. G. Blair, J. A. Bryant, M. J. Buckingham, J. A. Davidson, C. Edwards, J. Ferreirinho, K. D. Griffiths, R. N. James, F. van Kann, A. G. Mann, L. D. Mann, P. J. Veitch, GR9 Abstracts (ref. 46), Vol. 2, p. 411.
[22] G. V. Pallottino and G. Pizzella, GR9 Abstracts (ref. 46), Vol. 2, p. 420 (1980).
[23] E. Amaldi, I. Modena, G. V. Pallotino, G. Pizzella, F. Ricci, P. Bonifazi, F. Bordoni, F. Fuligni, U. Giovanardi, V. Iafolla, and S. Ugazio, GR9 Abstracts (ref. 46), Vol. 2, p. 410 (1980).
[24] D. H. Douglass, M. W. Cromar, R. Q. Gram, W. W. Johnson, C. Lam, and M. Macaluso, in: Proceedings of the Second Marcel Grossman Meeting on General Relativity, Ed. R. Ruffini (North Holland, 1980).
[25] J. Kadlec and W. O. Hamilton, in: Proceedings of the Second International Conference on Superconducting Quantum Devices, Berlin (1980), in press.
[26] V. I. Panov and F. Ya. Khalili, GR9 Abstracts (ref. 46), Vol. 2, p. 397 (1980).
[27] V. P. Mitrofonov and H. Lemann, GR9 Abstracts (ref. 46), Vol. 2, p. 418 (1980).
[28] Stanford Gravity-Wave Detector Group, "Preliminary Evaluation of the Performance of the Stanford 4800 kg Low-Temperature Gravitational Wave Detector", Stanford University Preprint (June 11, 1980).
[29] T. Suzuki, T. Tsubono, and H. Hirakawa, Phys. Lett. **A 67**, 2 (1978).

[30] H. Hirakawa, GR9 Abstracts (ref. 46), Vol. 2, p. 382 (1980); also H. Hirakawa, K. Tsubono, and M.-K. Fujimoto, Phys. Rev. **D 17**, 1919 (1978).
[31] R. L. Forward, Phys. Rev. **D 17**, 379 (1978).
[32] R. Weiss, Quarterly Prog. Rep., Research Lab. Electronics, MIT, **105**, 54 (1972).
[33] A. Rüdiger, R. Schilling, L. Schnupp, W. Winkler, H. Billing, and K. Maischberger, preprint MPI-PAE/Astro 233 (Max Planck Institut, Munich, 1980).
[34] R. Schilling, L. Schnupp, W. Winkler, H. Billing, K. Maischberger, and A. Rüdiger, preprint MPI-PAE/Astro 231 (Max Planck Institut, Munich, 1980).
[35] R. W. P. Drever et al., in: Proceedings of the Second Marcel Grossman Meeting on General Relativity, Ed. R. Ruffini (North Holland, 1980).
[36] J. W. Armstrong, R. Woo, and F. B. Estabrook, Astrophys. J., **230**, 570 (1979).
[37] R. W. Hellings, GR9 Abstracts (ref. 46), Vol. 2., p. 380 (1980).
[38] R. W. Hellings, J. D. Anderson, P. S. Callahan, and A. T. Moffet, Phys. Rev. **D**, submitted.
[39] J. Weber, Phys. Rev. Lett., **18**, 498 (1967).
[40] A. J. Anderson, private communication.
[41] L. L. Smarr, R. F. C. Vessot, C. A. Lundquist, and R. Decher, J. Gen. Rel. Grav., in press.
[42] J. Turneaure, private communication.
[43] V. B. Braginsky and S. P. Vyatchanin, Zhur. Eksp. Teoret. Fiz., **74**, 828 (1978); English translation in Soviet Physics-JETP, **47**, 433 (1978); see also reference 7.
[44] R. Weiss, in: Sources of Gravitational Radiation, Ed. L. Smarr (Cambridge University Press, 1979).
[45] R. Decher, J. Randall, P. Bender, and J. Faller, "Design Aspects of a Laser Gravitational Wave Detector in Space", preprint, Marshall Spaceflight Center (Huntsville Alabama) and Joint Institute for Laboratory Astrophysics (Boulder, Colorado).
[46] GR9 Abstracts: Abstracts of Contributed Papers for the Discussion Groups, 9th International Conference of General Relativity and Gravitation, July 14—19, 1980, Friedrich Schiller University, Jena, DDR.

Address: Prof. Dr. Vladimir B. Braginsky
Moscow State University
Moscow 117234
USSR

Prof. Dr. K. S. Thorne
California Institute of Technology
Dept. of Physics
Pasadena, CA 91125
USA

Experiments on Gravitational Waves with Electromagnetic Detectors

L. P. Grishchuk (Moscow)

1. Introduction

The first efforts at detecting gravitational waves involved the use of mechanical systems like metal bars, crystals, etc. There is substantial recent progress, both theoretical and experimental, along this line, as we have learned from the lecture by Braginsky and Thorne. In view of this, why are we interested in the interaction of gravitational waves with electromagnetic fields or, more pragmatically, why are we concerned with electromagnetic detectors of gravitational waves? Is there anything special in the interaction of gravitational waves with electromagnetic fields which distinguishes it from interaction with mechanical systems? Yes, there is. Let us mention two features. The first is the fact that the velocities of propagation of electromagnetic and gravitational waves are equal. As a result of this many coherent effects become possible. The second is that the kind of influence which a gravitational wave exerts on a mechanical oscillator, such as a metal bar, is very different from that exerted on an electromagnetic oscillator, such as a standing electromagnetic wave in a cavity. In the first case the gravitational wave acts as an external force while in the second case it exerts a parametric influence. This leads to important differences between the interactions of the gravitational wave with the two oscillators on both the classical and quantum levels. These differences create a certain hope that electromagnetic devices might become effective and fundamentally new tools for investigating gravitational radiation.

2. The geometric optics approximation

Einstein was probably the first to invent electromagnetic detectors of gravitational fields. He discussed the frequency shift and angular deviation of light rays in the gravitational field of a heavy body, which are examples of electromagnetic probing of gravity. It is clear that in the field of a gravitational wave one will encounter similar effects though they will have some kind of periodic behaviour.

Let us consider a weak monochromatic gravitational wave [1, 2]

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - \mathrm{d}x^2 - (1 - a)\,\mathrm{d}y^2 - (1 + a)\,\mathrm{d}z^2 + 2b\,\mathrm{d}y\,\mathrm{d}z \tag{1}$$

where

$$a = h_+ \sin\,[q(x^0 - x) + \psi], \qquad b = h_x \cos\,[q(x^0 - x) + \psi].$$

In the limit of very high frequency the world line of a photon can be described by a null geodesic. By solving the geodesic equation in the space time (1) it is easy to show that the photon experiences periodic deviations of frequency and of propagation direction along its trajectory. The maximal deviations occur when the photon has propagated over a distance of the order of the gravitational wavelength $\lambda = 2\pi/q$. Their orders of magnitude are $\Delta\omega/\omega \sim h$, $\Delta l/\lambda \sim h$, where $h = \sqrt{h_+^2 + h_x^2}$ is the dimensionless amplitude of the gravitational wave.

It is instructive to see how a photon gets frequency shifted with respect to the local inertial frame of reference. Such a frame is the closest thing to the global inertial coördinate system which is normally used in flat space-time. Let a local inertial frame be constructed along the time-like geodesic $x = y = z = 0$. The frame is valid in the vicinity of this line, for distances $l \ll \lambda$. A nearby time-like geodesic describes the motion of a neighbouring particle under the action of the "Newtonian" gravitational potential $\varphi/c^2 \sim h(l/\lambda)^2$ of the gravitational wave. With respect to the local inertial frame the particle moves along an elliptical trajectory with characteristic velocity $v/c \sim h(l/\lambda)$ [3, 4]. Suppose that a photon was emitted from the origin of the local inertial frame. After travelling a distance l it acquires a frequency shift $(\Delta\omega/\omega)_g \sim \varphi/c^2 \sim h(l/\lambda)^2$ as measured with respect to this frame. If the frequency shift is measured with respect to a freely moving neighbouring particle or, in particular, if the photon is reflected by such a particle and returned, then the frequency shift is determined by $(\Delta\omega/\omega)_D \sim v/c \sim h(l/\lambda)$. In this case the main contribution to the frequency shift is not of "gravitational" but rather of "Doppler" nature. This simple example shows that in certain situations the photon's frequency shift serves as a sensing system for the motion in the field of a gravitational wave and not as an "electromagnetic detector" of the wave. If the distance of travel l is of the order of λ, then the two contributions are of the same order of magnitude.

Measurement of the frequency shift as a tool for detecting gravitational radiation was suggested in many works [5—8]. This effect lies at the foundation of a variety of methods. Among them are Doppler tracking of spacecraft [9—12], laser interferometry [13—14], angular anisotropy of the temperature of the cosmic microwave background [15—17], variation of the arrival time of signals from pulsars [18—20], etc.

In the field of a monochromatic gravitational wave the frequency shift can be accumulated if the photon is forced to move along a specific trajectory again and again. Some resonance condition between the frequency of the gravitational wave and the frequency of the photon's revolution needs to be satisfied. The first example of this kind is a pair of photons in a circular waveguide [6]. Under the condition $\Omega = 2\nu$, where Ω is the frequency of the gravitational wave and ν is the frequency of the photon's orbit, the frequency shift between two photons will linearly increase with time $\Delta\omega/\omega \sim h\Omega t$. Another example is a set of freely "floating" mirrors reflecting a photon and returning it to the initial position [21]. In fact it is sufficient to have only two freely moving mirrors properly oriented and separated by a distance $l = \lambda/2$ [7]. After Q reflections the frequency shift reaches $\Delta\omega/\omega \sim hQ$.

Small deviations in the direction of propagation can also be accumulated if the photon is reflected between two mirrors [7]. There will be systematic drift of the place where the photon hits a mirror. After Q reflections the position of this impact will move by a distance of order Δx, where $\Delta x/\lambda \sim hQ$.

3. Maxwell's equations in the field of a gravitational wave

The geometric optics approximation has a limited range of validity. If the space occupied by the electromagnetic field is large and (or) the scale of variability of the electromagnetic field is comparable with that of the gravitational wave, the full set of Maxwell's equations should be analyzed. In curved space-time the Maxwell equations have the form

$$F^{\alpha\beta}{}_{;\beta} = -\frac{4\pi}{c} j^{\alpha}, \qquad F_{\alpha\beta,\gamma} + F_{\gamma\alpha,\beta} + F_{\beta\gamma,\alpha} = 0. \tag{2}$$

The interaction of electromagnetic fields with gravitational waves was considered in many works [22—28]. The first approaches to this problem were a little bit naive. They did not take properly into account the boundary conditions, they did not analyse the most favorable configurations of the electromagnetic field, they unrealistically assumed that the boundaries are penetrable for electromagnetic radiation, they ignored some of the resonant effects, and so on. However, two very important things were learnt: gravitational waves can convert into electromagnetic radiation; and conversion of the waves can be coherent, i.e. lasting for many periods of the waves.

To obtain an idea of how a gravitational wave can convert into an electromagnetic one, let us consider a volume of space with characteristic dimension l occupied by a constant magnetic field H. The gravitational wave enters this volume and, roughly speaking, begins to displace the magnetic lines of force producing a variable electromagnetic field. This perturbation propagates with the gravitational wave at the velocity of light. However, such a perturbation is nothing other than an elctromagnetic wave. The amplitude of this wave is rising as the gravitational wave propagates further and further. The transformation coefficient of the gravitational flux density W_g into electromagnetic flux density W_e is

$$k = \frac{W_g}{W_e} \sim \frac{GH^2}{c^4} l^2 \sim \frac{r_g}{l},$$

where r_g is the gravitational radius of the total mass of the magnetic field in the volume.

The opposite process, transformation of an electromagnetic wave into a gravitational one, also takes place. The two waves are coupled via the constant magnetic field and constitute a united physical system. The normal modes of this system represent combinations of electromagnetic and gravitational waves and can be called "gravi-photons" [29, 30]. Despite the fact that only under very special physical conditions can complete mutual conversion of the waves be achieved, this effect is important for understanding more practical processes such as the interaction of gravitational waves with electromagnetic fields in a cavity.

A rigorous treatment of the influence of a gravitational wave on the electromagnetic field in an resonator with the boundary conditions properly taken into account was started in [31], see also [32, 33]. By decomposing the electromagnetic field into normal modes one can cast Eq. (2) into a Hamiltonian form. If the magnetic field is associated with the generalized coordinate q and the electric field with the generalized momentum p, then Eq. (1) can be represented as [34]:

$$\dot{q}_m = p_m + \sum_n p_n C_{mn}(t), \qquad -\dot{p}_m = \omega_m^2 q_m + \sum_n \omega_n^2 q_n A_{mn}(t) + B_m(t). \tag{3}$$

Here indices m, n denote the number of the modes. The coefficients A_{mn}, B_m, C_{mn} describe the coupling of the electromagnetic field to the wave. They depend on the shape of the resonator, its orientation, polarization of the wave, etc. Eq. (3) can be derived from a Hamiltonian. Among the systems described by (3) the one-mode system has the simplest Hamiltonian

$$H = \frac{1}{2}\left(1 + C(t)\right)p^2 + \frac{1}{2}\,\omega^2\left(1 + A(t)\right)q^2 + B(t)\,q.$$

The last term describes the action of the external force and is only present if a constant electromagnetic field exists in the resonator initially. The first two terms exhibit the parametric action of the wave. They change the frequency and the "mass" of the electromagnetic oscillator.

The resonant effects occur if $\omega_m \pm \omega_n = \Omega$, where Ω is the frequency of the gravitational wave. The variations produced by the wave can be accumulated over the time $\tau^* \sim Q/\omega$, where Q is the quality factor of the resonator. Three typical cases should be mentioned. a) A constant field is present in the resonator initially ($\omega_n = 0$, $\omega_m = \Omega$). Finally, the normal mode ω_m will be excited. The total energy $\Delta\varepsilon$ in the lowest mode is related to the total energy ε of the field in the resonator by the formula $\Delta\varepsilon/\varepsilon \sim (hQ)^2$. b) A normal mode ω is present initially (parametric amplifier, $\Omega = 2\omega$). Under certain phase conditions the energy of the mode or the phase of the oscillations will be changed after time τ^* by $\Delta\varepsilon/\varepsilon \sim hQ$ or $\Delta\varphi/\varphi \sim hQ$. c) Two normal modes are excited initially (parametric convertor, $\omega_m - \omega_n = \Omega$). The energy will be transferred from one mode to the other according to the relation $(\Delta\varepsilon/\varepsilon) \sim (hQ)^2$. An important class of "tunable" detectors belongs to this case [35, 36].

It is worthwhile giving an estimate of the possible sensitivity of these electromagnetic detectors. Let the frequency of the gravitational wave be $\Omega \approx 10^9$ rad/sec. Half of the wavelength is $\lambda/2 \approx 10^2$ cm, which determines the characteristic size of the simplest detector of type a). For $H \approx 10^5$ Gauss and $Q \approx 10^{12}$ one gets fairly good sensitivity: $h \approx 10^{-27}$, $I \approx 10$ erg/sec cm^2. This sensitivity was derived at the level of the so-called "standard quantum limit" (for definition, see [37, 38]). Application of the ideas of "quantum nondemolition measurements" (see [38, 39] and below) can substantially improve this sensitivity. Unfortunately periodic astrophysical sources in this frequency band are not known. To observe the existing low frequency astrophysical sources one should exploit type c) detectors whose sensitivity is probably worse.

4. Laboratory gravitational wave experiment

An electromagnetic field can serve not only as a detector but also as an effective source of gravitational radiation [40, 41]. Let us compare the performances of mechanical and electromagnetic emitters. For the comparison to be fair we assume that they radiate gravitational waves at the same frequency $\Omega = (2\pi c)/\lambda$ and their performances are compared at equal distance from the emitters. A model for a mechanical emitter (m-emitter) is a piece of vibrating solid-state material and a model for an electromagnetic emitter (e-emitter) is a standing electromagnetic wave in a cavity. We start by considering elementary m- and e-emitters, which means that they oscillate in the lowest normal mode and have sizes of order of the acoustic wavelength $\lambda_s \approx (v_s/c)\lambda$ and the electromagnetic wavelength $\lambda_e \approx \lambda$, respectively.

The characteristic amplitude h of the gravitational wave at the distance r is determined by

$$h_{ik} \approx \frac{G}{c^4}\frac{1}{r}\int T_{ik}\,\mathrm{d}V .$$

Suppose A is the amplitude of elastic vibration in the elementary m-emitter. Then the stress tensor is of order of $\sigma_m \sim \varrho_m v_s^2(A/\lambda_s)$. As the result one obtains $h_m \sim \tilde{r}_g/r(v_s/c)^2$, where $\tilde{r}_g$ is the gravitational radius of the variable part of the mass of the emitter:

$$\tilde{r}_g \approx \frac{G\varrho_m\lambda_s^3}{c^2}\cdot\frac{A}{\lambda_s}.$$

Similarly for the elementary e-emitter one obtains

$$h_e \sim \frac{r_g}{r},$$

where

$$r_g \approx \frac{G\varrho_e\lambda_e^3}{c^2}.$$

For moderate values of the parameters appearing in these relations the elementary e-emitter is much more effective than the elementary m-emitter. For example, if $\varrho_m \approx 1\ \mathrm{g/cm^3}$, $\varrho_e \approx 10^{-18}\ \mathrm{g/cm^3}$, $v_s/c \approx 10^{-5}$, $A/\lambda_s \approx 10^{-3}$, one obtains $h_m/h_e \approx 10^{-10}$. However, the volume of the elementary e-emitter, λ_e^3, is much larger than the volume of the elementary m-emitter, λ_s^3. The former can contain $N = (\lambda_e/\lambda_s)^3 \approx (c/v_s)^3 \gg 1$ elementary m-emitters. One may ask for the efficiency of the m-emitter with the same total volume, λ_e^3, as the e-emitter. Here an advantage of the electromagnetic sources manifests itself. Indeed, the e-emitter of this volume is automatically coherent, radiation from different parts of the volume does not interfere destructively. While to achieve the coherence of N elementary m-emitters one must make them specially phased. Suppose this was somehow realized. Then the amplitude of the gravitational wave from such a source is $h_{mc} \approx Nh_m$. Hence,

$$\frac{h_{mc}}{h_e} \approx \frac{\varrho_m}{\varrho_e}\left(\frac{v_s}{c}\right)^2\frac{A}{\lambda_s} \approx \frac{\sigma_m}{\sigma_e}.$$

So, after all, the efficiency of the emitters is determined by the magnitude of the feasible variable stresses provided that the m-emitter is made coherent. One can, probably, obtain electromagnetic stresses as large as the highest possible mechanical variable stresses for normal materials by producing and maintaining electromagnetic fields with feasible strength.

A complete proposal for emitting and detecting gravitational waves by electromagnetic resonators in laboratory conditions was suggested in [33]. The proposed scheme takes into account the best geometrical factors, favourable orientation of the resonators and so on. It was shown that the signal to noise ratio approaches unity if the total volume of the system is $V \approx 25 \cdot 10^9$ cm^3, electromagnetic field strength $H \approx 3 \cdot 10^5$ Gauss, quality factor of the resonator-detector $Q \approx 7 \cdot 10^{13}$. These figures illustrate the enormous efforts which have to be undertaken in order to perform this experiment, which is analogous to the Hertz experiment in electrodynamics. However, it should be emphasized that the detectability condition in this estimate was accepted at the level of the "standard quantum limit". The experiment may become more realistic if one is fortunate enough to apply "quantum nondemolition" measurements.

5. Quantum theory of electromagnetic detectors and quantum nondemolition measurements

To detect gravitational waves from realistic sources one will probably need such a high sensitivity that the quantum properties of the (macroscopic) gravitational detectors may become important.

Hamiltonian form of Maxwell's equations, derived in Sect. 3, provides a basis for the quantum theory of electromagnetic oscillators [34]. The simplified Hamiltonian for the one-mode system is

$$\hat{H} = \frac{1}{2}\,\hat{p}^2 + \frac{1}{2}\,\omega^2\big(1 + A(t)\big)\,\hat{q}^2,$$

where $\hat{p}$ and $\hat{q}$ are operators which satisfy the usual commutation relations:

$$[\hat{p}, \hat{p}] = 0 = [\hat{q}, \hat{q}], \qquad [\hat{q}, \hat{p}] = i\hbar.$$

This Hamiltonian manifestly describes perturbation of the frequency induced by a gravitational signal. (For different approaches to the problem see [42, 43]). Hamiltonians of that kind are well known in nonrelativistic quantum mechanics (see, for example, [44]). The parametric nature of the perturbation results in different transition probabilities as compared with the case of a force acting on the quantum oscillator.

Let the oscillator be initially in the n-quantum state. The largest probabilities of changing this state under the action of the external force are $P_{n,n+1} \sim h^2(n+1)$ and $P_{n,n-1} \sim h^2 n$. For the case of the parametric influence the largest probabilities are $P_{n,n+2} \sim h^2(n+1)(n+2)$, and $P_{n,n-2} \sim h^2 n(n-1)$. Suppose an experimenter is able to detect a gravitational signal so weak that it causes a transition to one of the nearest levels. Then, for large n, the minimal detectable h is of order of $h_{\min} \sim 1/\sqrt{n}$ in the first case and $h_{\min} \sim 1/n$ in the second case. The different and more favourable dependence of $h_{\min}$ on n gives a certain advantage to the electromagnetic detector over the mechanical one.

All the classical formulae, expounded in Sect. 3, can certainly be obtained in a straightforward quantum mechanical way. So, one may admit, indeed, that "sometimes quantum mechanics helps us to understand classical mechanics" [45].

In the attempt to reach the highest possible sensitivity one of the most intriguing questions is the possibility of avoiding restrictions imposed by the quantum-mechanical uncertainty principle. Measurements which do not pretend to know simultaneously the precise values of noncommuting observables can have sensitivities better than the "standard quantum limit". To perform such a measurement one needs to know the quantum nondemolition (QND) operators (observables). A Heisenberg picture operator $Z(t)$ which continuously depends on time is called a QND operator if it satisfies the commutation relation

$$[Z(t), Z(t')] = 0$$

for any t and t' [46, 38]. QND operators for measuring force (QNDF operators) are known [38, 47]. QND operators for measuring a parametric influence (QNDP operators) were introduced in [34]. One can distinguish QND operators of two classes. The first class, which can be called a class of "simultaneous" QND operators, is characterized by the demand that $Z(t)$ is constructed from $\hat{p}(t)$ and $\hat{q}(t)$ taken at the same moment of time. The second class can be called a class of "shifted" operators and is characterized by the use of $\hat{p}$ and $\hat{q}$ taken both at t and at previous moments of time $t - \tau$, to construct $Z(t)$.

It was explicitly shown that QNDP operators of both classes do exist [34]. For measuring the operators (observables) of the first class one needs complete knowledge of the time dependence of the gravitational signal. So this class of QNDP measurements can be applied to detecting gravitational waves from double stars, in a laboratory experiment, and so on. Observables of the second class are independent of any a priori knowledge whatsoever about the time-structure of a gravitational signal.

The existence of QNDP operators shows that, in principle, there is no limit to precise measurement of the parametric influence on the quantum oscillator. Together with other advantages mentioned above, this says that it is worthwhile constructing and using electromagnetic detectors in gravitational wave experiments.

I am indebted to M. MacCallum for a critical reading of the manuscript.

References

[1] C. W. Misner, R. S. Thorne, J. A. Wheeler, Gravitation. Freeman. San Francisco (1973).
[2] L. D. Landau, E. M. Lifshitz. The classical theory of fields. Pergamon Press, New York (1975).
[3] L. P. Grishchuk. Sov. Phys. Usp. **20**, 319 (1977).
[4] L. P. Grishchuk, A. G. Polnarev. General Relativity and Gravitation (Einstein Century Volume) ed. A. Held, v. II, 393 (1980).
[5] W. J. Kaufmann. Nature **277**, 157 (1970).
[6] V. B. Braginskii, M. B. Menskii. JETP Lett. **13**, 417 (1971).
[7] L. P. Grishchuk. Sov. Phys. JETP **39**, 402 (1974).
[8] W. L. Burke. Astrophys. J. **196**, 324 (1975).
[9] F. B. Estabrook, H. D. Wahlquist. GRG **6**, 439 (1975).
[10] A. J. Anderson, in Experimental Gravitation, ed. B. Bertotti (New York, Academic Press), p. 163, (1974).
[11] B. Mashhoon, L. P. Grishchuk. Astrophys. J. **236**, 990 (1980).
[12] R. W. Hollings. Spacecraft-Doppler gravity wave detection, J.P.L. preprint (July, 1980).
[13] R. Weiss. Quart-progress report. M.I.T. **105**, 54 (1972).
[14] R. L. Forward. Phys. Rev. **D 17**, 379 (1978).
[15] R. K. Sachs, A. M. Wolfe. Astrophys. J. **147**, 80 (1967).
[16] G. Dautcourt, in Confrontation of Cosmological Theories with Observational Data, ed. M. Longair, p. 299, Reidel (1974).
[17] L. P. Grishchuk, Ya. B. Zeldovich. Astronomicheskii Zhurn. **55**, 209 (1978).
[18] M. V. Sazhin. Soviet. Astron. A. J. **22**, 36 (1968).
[19] J. S. Detweiler. Astrophys. J. **234**, 1100 (1979).
[20] B. Mashhoon. Caltech preprint OAP-594 (July 1980).
[21] L. Halpern. Bull. cl. sc. Acad. Roy. Belg. **58**, 647 (1972).
[22] M. F. Gertsenshtein. Sov. Phys. JETP **14**, 84 (1962).
[23] G. A. Lupanov. Sov. Phys. JETP **25**, 76 (1967).
[24] F. Cooperstock. Ann. Phys. (N.Y.) **47**, 173 (1968).
[25] H. Heintzman. Zs. Phys. **210**, 380 (1968).
[26] D. Bocaletti, V. De Sabbata, P. Fortini, C. Gualdi. Nuovo Cimento, Ser. X, **70B**, 129 (1970).
[27] D. Bocaletti, F. Occhionero. Lett. Nuovo Cimento **2**, 549 (1971).
[28] V. L. Ginsburg, V. C. Tzitovich. Izvestia VUZov Radiofizika **18**, 173 (1975).
[29] Ya. B. Zeldovich. Sov. Phys. JETP **38**, 652 (1974).
[30] I. N. Gerlach. Phys. Rev. Lett. **32**, 1023 (1974).
[31] V. B. Braginskii, L. P. Grishchuk, A. G. Doroshkevich, Ya. B. Zeldovich, I. D. Novikov, M. V. Sazhin. Sov. Phys. JETP **38**, 865 (1974).
[32] P. Tourrenc, J. L. Crossiord. Nuovo Cimenteo **19B**, 105 (1974).
[33] L. P. Grishchuk, M. V. Sazhin. Sov. Phys. JETP **41**, 787 (1975).
[34] L. P. Grishchuk, M. V. Sazhin. Sov. Phys. JETP.
[35] C. M. Caves. Phys. Lett. **80B**, 323 (1979).
[36] F. Pegoraro, E. Picasso, L. Radicati. J. Phys. **A 11**, 1949 (1978).
[37] V. Braginskii, A. Manukin. Measurements of Weak Forces in Physics Experiments. University of Chicago Press, Chicago (1977).
[38] C. M. Caves, K. S. Thorne, R. Drever, V. Sandberg, M. Zimmerman. Rev. Mod. Phys. **52**, 341 (1980).
[39] V. B. Braginskii, Yu. I. Vorontsov, F. Ya. Khalli. Sov. Phys. JETP **46**, 705 (1977).
[40] L. P. Grishchuk. Phys. Lett. **56A**, 255 (1976).
[41] L. P. Grishchuk, M. V. Sazhin. Sov. Phys. JETP **38**, 215 (1974).
[42] N. Grafe, H. Dehnen. Intern. Journ. Theor. Phys. **15**, 393 (1976).

[43] P. Tourrenc. G.R.G. **9**, 123—140 (1978).
[44] A. I. Baz, Ya. B. Zeldovich, A. M. Perelomov. The Scattering, Reactions and Fissions in Nonrelativistic Quantum Mechanics, Nauka, Moscow (1971).
[45] P. Paradoxov. Uspekhi Fiz. Nauk **89**, 707 (1966).
[46] W. G. Unruh. Phys. Rev. **D 19**, 2888 (1979).
[47] V. V. Dodonov, V. I. Man'ko, V. N. Rudenko. Zhurn. Exper. Teoret. Fiz. 78, 88 (1980). Sov. Phys. JETP (1980).

Address: Dr. L. P. Grishchuk
Shternberg Astron. Institute
117234 Moscow
USSR

A Gravity-Wave Detector Using Optical Cavity Sensing[1])

R. W. P. Drever, G. M. Ford, J. Hough, I. M. Kerr, A. J. Munley, J. R. Pugh, N. A. Robertson, and H. Ward (Glasgow)

Most of the current experimental work on development of gravitational wave detectors with optical sensing is based on use of Michelson interferometers to monitor relative separations between test masses [1]. Multiple reflection mirror systems are employed to cause the light beams to traverse the distance between the test masses a large, but discrete, number of times and enhance displacement sensitivity. An obvious practical difficulty with this arrangement is that for separations of the order of a kilometre between the masses the size of the mirrors (and of the vacuum pipes) required becomes large. A less obvious, but serious, problem is that incoherent scattering at the mirrors can give rise to significant noise unless special precautions are taken [2]. It is well known that for many purposes a Fabry-Perot cavity can replace a Michelson interferometer, and the idea that a gravity wave detector might be made using a Fabry-Perot cavity formed between mirrors attached to a pair of free test masses is not new [3]. However there are very evident practical difficulties. If one were to attempt to look for the change in transmission of the optical cavity induced by the apparent motions of the end masses caused by the gravity wave the light source used would require to have fractional frequency stability of the order of the gravity wave amplitude — many orders better than that of present lasers. The source requirements might be reduced by using a comparison between transmissions of two cavities with their axes perpendicular to one another. However with optical cavities of length of order one kilometre, and of the high finesse desirable, the width of the cavity transmission peaks would be so small that for efficient use of the light the laser would have to possess exceptional stability and be locked to the cavities with unusual precision.

Standard methods of locking lasers to stable cavities usually involve monitoring the change of transmission of the cavity when the laser frequency changes and feeding back a control signal to the laser. The bandwidth of such a control system is limited by the bandwidth of the cavity, and would be quite inadequate for the present purpose. We have therefore proposed a method of locking a laser to a cavity in which a comparison is made between the phase of light from the laser with the phase of the stored light within the cavity. In practice this may be done by phase modulating the laser beam at a high radiofrequency, feeding the light to the cavity, and detecting the light reflected back from the input mirror of the cavity. The output from the photodetector passes through a phase sensitive detector referenced to the radio-

---

1) Abstract, submitted too late.

frequency modulation source, and then is applied to a Pockels cell phase modulator within the laser cavity itself, which provides a fine control of the laser frequency. Analysis of this arrangement shows that the control signal obtained corresponds to the phase difference between the light from the laser and the light emerging from the cavity, so that the bandwidth of the feedback loop is not in any way limited by the storage time of the cavity itself and it seems practicable to lock the laser sufficiently well for at least initial tests of an optical cavity gravity wave detector.

It may be noted that the laser locking system just described is a fairly close optical analogue of a technique used to stabilise microwave oscillators, originated by R. V. Pound in 1944.

The gravity wave detection system which we propose has many possible variants. In one simple form there are two optical cavities perpendicular to one another, with the mirrors attached to three suspended test masses. Light from an argon laser passes through a beamsplitter into both of the cavities so formed.

The laser light is phase modulated, and the reflected light from one cavity is used to stabilise the laser. A similar phase measuring arrangement is then used to adjust the distance between the masses forming the second cavity so that this cavity becomes locked to the laser. Gravity wave signals may then be looked for by examining the residual phase errors in the two control loops and the feedback forces or displacements applied to the second cavity.

The above arrangement, although simple in concept, would not be expected to give optimum signal to noise ratio. Better performance could be obtained by making the measurement of phase difference between the two cavities a direct optical one: recombining the light emerging from the cavities at a second beamsplitter and possibly using high frequency differential phase modulation of the two beams before recombining to remove low frequency laser amplitude noise. Again, a symmetrical arrangement, in which the laser is locked to the mean of the two cavities, has advantages. Indeed there are many variants of the basic system which we have proposed.

Precise matching of the phase response of the two cavities is important to reduce the extremely high performance demanded of the laser locking system, and subsidiary control loops may be necessary to achieve this. In practice the complete system is likely to become significantly more complex than a gravity wave detector based on a Michelson interferometer.

In spite of the obvious difficulties and complexity of this proposed gravity wave detector we have felt it worth investigating experimentally. As a first stage we have used the laser stabilising system to phase lock an argon ion laser at Glasgow to a 10-metre optical cavity. Some parallel experiments have also been carried out in collaboration with J. L. Hall and F. W. Kowalski at the Joint Institute for Laboratory Astrophysics, University of Colorado, to more precisely investigate the phase locking performance of the stabilising scheme. Results so far are encouraging, but a considerable amount of work will be necessary before it will be possible to assess the real practicability of the scheme proposed as a gravitational wave detector.

References

[1] Drever, R. W. P. et al. Proc. Roy. Soc. **A 368**, 11 (1979); Proceedings of the International Meeting on Experimental Gravitation, Pavia, Italy, 1977, p. 365.

[2] Hough, J. et al. This Meeting ("Discussion Group on Free Mass Interferometer Antenna"). Drever, R. W. P. et al.; Maischberger et al.; The second Marcel Grossmann Meeting on the Recent Developments of General Relativity, Trieste, July 1979.
Billing et al. J. Phys. E., Sci. Instrum., 12, 1043—50, (1979).

[3] See, for example, Braginsky, V. B. and Manukin, A. B. "Measurement of Small Forces in Physical Experiments" (Nauka, Moscow, 1974) (English translation: University of Chicago Press, 1977).

Address: Prof. Dr. R. W. P. Drever
Dept. of Natural Philosophy
Glasgow University
Glasgow G12 8QQ
Scotland (U.K.)

HISTORICAL PAPERS

Some Supplements to Einstein-Documents

H. Melcher (Erfurt)

Preliminary remarks

In the series of International GR Conferences, a discussion group on "Historical Aspects" was first established at GR9 in Jena. It found a lively response. The extensive programme took more time than would normally have been available.

There were four lectures:

1. Lewis Pyenson (Montreal): Physical sense in relativity: Max Planck edits the Annalen der Physik, 1906—1918.
2. John Stachel (Princeton): Einstein's struggle with general covariance, 1912—1915.
3. Liviu Sofonea (Brașov): Some relativistic ideas in the prerelativistic physics.
4. Rainer Schimming (Leipzig): Historical sketch of gravitational waves.

Unfortunately all these contributions cannot be published in this volume.

The GR 9 Conference in Jena came at the end of the Einstein centenary: J. A. Wheeler introduced the congress with his lecture "Einstein's second century". The theme of the following contribution was chosen with this in mind. It may also be understood as gleanings at the end of a number of celebrations.

Einstein and the German Physical Society [1]

Until now Einstein's work in the German Physical Society is rarely mentioned in the Einstein literature — and only briefly. These activities have claimed a considerable part of Einstein's energies during his first 10 years in Berlin.

When Einstein came to Berlin in April 1914, he was already a "foreign member" of the German Physical Society. In addition to many renowned physicists from Europe and the rest of the world, institutes, factories, firms, libraries, and societies from outside Germany were also members — for example the Physical Society (Lebedev Society) in Moscow.

Einstein's affiliation as a foreign member (Zürich, Hofstr. 116), dated 7. 11. 1913, was proposed by Max von Laue. After he moved to Berlin, Einstein's name appeared as a Berlin member of this society (first Wittelsbacherstr. 13, then from 1918 until his emigration: Haberlandstr. 5).

The board of directors of the society was regularly elected in May. On 8. 5. 1914 Einstein became a member („Beisitzer", assessor) of this board (chairman: Fritz

Haber). He was re-elected as an assessor for a second annual term on 14. 5. 1915 (chairman: Max Planck). Albert Einstein was then elected chairman of the German Physical Society on 5. 5. 1916 and again on 11. 5. 1917. In this function he was followed by Max Wien (10. 5. 1918) and Arnold Sommerfeld (31. 5. 1918). Einstein belonged to the board of directors until 1925. He had had to work especially hard during the complicated period after World War I, e.g. on working out a new statute and above all reorganizing the physical journals.

The main source of information about the activities of Einstein as a member, assessor and chairman of the German Physical Society is the journal of this society: Verhandlungen der Deutschen Physikalischen Gesellschaft. Here one finds a piece of history of physics, accurately handed down by Karl Scheel. His name is connected for all time with physical literature.

Original minutes or other handwritten drawings or reports of colleagues, with further details for the history of sciene about Einstein's activities in the German Physical Society, are so far unknown.

One first finds Einstein's name in 1909 in the journal of the society (p. 417), in the programme of the 81. Versammlung Deutscher Naturforscher und Ärzte in Salzburg. There Einstein gave his famous lecture on 21. 9. 1909: "On the development of our views on the nature and constitution of radiation". This lecture was published in the journal of the German Physical Society and in the Physical Journal [2]. Einstein's name appeared again four years later when his report "On the gravitational problem" was announced for the 85. Versammlung Deutscher Naturforscher und Ärzte in Vienna (23. 9. 1913).

Before his membership in the German Physical Society Einstein's name had also appeared in its journal when his friend Paul Habicht delivered a lecture in the session of the society on 15. 12. 1911 (chairman: Fritz Kurlbaum). The journal mentioned: "Hr. Ing. Habicht demonstrated Einstein's potential-multiplier".

Einstein had stimulated the construction of this apparatus with his article in 1908 "A new electrostatic method for measuring small electrical charges". There he mentioned that it would be important to improve the sensitivity of electrostatic methods for research in radioactivity [3]. Two years later Conrad and Paul Habicht from Schaffhausen (Switzerland) published in the same journal the article "Elektrostatischer Potentialmultiplikator nach A. Einstein" [4]. Here they reported that they had carried out the experiments together with Einstein in the university-laboratory in Zürich and that they had applied for a patent for the construction. Unfortunately it is not clear whether Einstein was named as a co-inventor. Nothing is known about the correspondence with the patent office and if it was patented or not.

Einstein was actively interested in the development of this small electrostatic machine ("elektrostatisches Maschinchen") and wrote about it several times in letters to his friend Michele Besso [5]. So he wrote from Prague (26. 11. 1911) that Habicht had demonstrated his small machine to the Physical Society in Berlin with "enormous success" and that it was superior to the string-electrometer. In another letter from Prague on 4. 2. 1912 he wrote once more to Besso: "I have noted with certainty the success he had in Berlin; the fellows almost stood there on their heads".

In 1920 Walther Gerlach measured the contact potentials with such an apparatus in Tübingen [6]. It was still demonstrated in 1979 in an exhibition in honour to Einstein, v. Laue, Hahn and Meitner [7]. This example carried a small plate: Paul Habicht, Schaffhausen, Potentialmultiplikator Typ 6 Nr. 18, Motor 1/150 PS, Volt 2—6, plombiert.

Einstein also wrote about this "Maschinchen" to Albert Gockel (3. 12. 1908). It was his aim to proof the validity of the molecular theory (Brownian movement) in the fields of electrostatics [8].

By the end of the year 1914 Einstein had already delivered three lectures to the Physical Society. Altogether he gave 23 lectures, 11 in his two years as the chairman of the society. Eighteen lectures were published, six of them in the journal of the society, others in „Naturwissenschaften", edited by Einstein's friend, Arnold Berliner, and others in the „Sitzungsberichte der Preußischen Akademie der Wissenschaften". There are also some articles by Einstein in the „Verhandlungen" on subjects which were not reported to the sessions of the society. The large number of lectures given by Einstein to the Physical Society during the time of his membership can scarcely be matched by any other member.

To the pioneer works that were reported to the German Physical Society belong very important lectures which have a place of honours in its history. Usually the sessions took place in the large lecture-theatre of the Physical Institute of the University (Am Reichstagsufer 7/8) [9]. Some of these outstanding lectures may be mentioned: Hermann v. Helmholtz (23. 7. 1847) "On the conservation of force"; Max Planck (14. 12. 1900) "On the theory of the law of energy-distribution in the normal-spectrum", which introduced the quantum theory.

The sessions of the society promoted knowledge about the laws of black radiation, as here were discussed mainly the works, experiments and measurements of W. Wien, O. Lummer, E. Pringsheim, H. Rubens, and F. Kurlbaum.

The famous experiments of J. Franck and G. Hertz, carried out in 1913 in the Physical Institute of the University (Director: F. Rubens) were also reported before the Physical Society.

Important chapters of the history of physics were written in Berlin. To the lectures which were important in the history of the Physical Society and of physics belongs the report of A. Einstein (19. 2. 1915) on the experiment carried out together with W. J. de Haas, first showing the existence of Ampère's molecular currents, later known as the Einstein-de Haas-Effect.

Einstein and de Haas worked as guests in the Physikalisch-Technische Reichsanstalt (President: Emil Warburg). Later, after 1916, Einstein belonged to the „Kuratorium" of this institution. Einstein did not report in the sessions of the Prussian Academy of Science on this gyromagnetic effect, found by him and de Haas, however apart from the session of the Physical Society in Berlin in the Academy of Science in Amsterdam.

Einstein wrote to Besso [10] (12. 2. 1915): "A wonderful experiment; it is a pity you cannot see it. And how crafty nature is, if one wants to get at her experimentally. I still get a passion for experimenting in my later days."

Remarkable for his role as a teacher is the fact that Einstein demonstrated the gyromagnetic effect — shortly after finishing his general theory of relativity — as an experiment for lecture halls [11].

When Einstein was chairman of the German Physical Society, the 60th birthday of Max Planck (23. 4. 1918) has been celebrated. Einstein had prepared and led this celebration. Besides the speeches of E. Warburg, M. v. Laue, and A. Sommerfeld Einstein's contribution "Motives of research" is an especially famous appreciation of the scientific personality of Max Planck; his assessment belongs to the classic texts of the theory and history of science [12].

On the occasion of the 70th birthday of Max Planck M. Born, A. Einstein, M. v. Laue, and A. Sommerfeld signed a call (at the end of 1927) for the establishment of a Max Planck medal [13]. On the day of Max Planck's Golden Doctor Jubilee (28. 6. 1929) Max Planck and Albert Einstein were awarded the first two medals.

Einstein led some sessions in Berlin too, which did not take place during his time as a chairman, especially in 1918/1919. Einstein led his first session of the society on 19. 5. 1915; after his chairmanship he led the one on 14. 3. 1919 (his 40th birthday), and his last session on 30. 7. 1920. Altogether he led 28 sessions. During the time of his serious illness (early in 1918) he was replaced by K. Scheel, H. Rubens and E. Warburg.

By tradition new members were admitted into the Physical Society only on the proposal of confirmed members. Albert Einstein proposed six persons for membership (1917 and 1918): Stud. phil. Rudolf Jakob Humm (Berlin); Privat-Doz. Dr. Fritz Noether (Karlsruhe); Dr. Wilhelm John (Berlin), Privat-Doz. Dr. Hans Thirring (Wien), Mrs. Dr. Jetty Cohn (Zürich) and Miss stud. math. Edith Einstein (Zürich).

Rudolf Jakob Humm from Aarau studied in Göttingen and visited Albert Einstein in Berlin in his bachelor-flat, asking Einstein to be introduced into the Physical Society. This first meeting with Einstein Humm described in his diary; it is recorded by Carl Seelig [14]. Later on Humm became a writer.

Fritz Noether was a brother of the famous Emmy Noether. He had to leave Nazi-Germany and was a professor of mathematics and mechanics in Tomsk (USSR).

Hans Thirring is well known with his work in the field of the theory of general relativity, particularly with the Thirring-Lense-effect. There were lifelong close relations between Einstein and Thirring.

Edith Einstein was a cousin of Albert Einstein. Her father, Jakob Einstein, was his uncle. Edith and Albert spent their young days together in Munich. Later Edith Einstein studied physics in Zürich and corresponded with Albert Einstein on scientific problems. She was married to the mathematician Reis and employed at a private school in Zürich [15].

In her thesis "On the theory of the radiometer" Edith Einstein thanked Albert Einstein for suggesting this work [16]. Einstein himself wrote an article about the "Theory of the radiometer-forces" [17] without reference to the work of Edith.

When Edith Einstein and Jetty Cohn became members of the Physical Society there were 304 members in Berlin and 459 foreign members, in each case including 7 women.

The sessions of the German Physical Society were occasions for meeting colleagues, indispensible for the development of science. Here Albert Einstein — for one example —

met Niels Bohr the first time. At the session of the Physical Society in Berlin on 27. 4. 1920 Niels Bohr gave a lecture on the "series of spectra of the elements". At this session Leo Szilard — at this time a student — became a member of the society. During their time in Berlin Szilard and Einstein received seven German patents (Deutsche Reichspatente) and six foreign patents [18]. Both men remained friends their entire lives.

As guests, further close friends of Einstein's were reported in the Physical Society: David Reichinstein (19. 10. 1924) in Prague and Rudolf Goldschmidt (18. 9. 1929) in Berlin. Later on Einstein declined to endorse the manuscript of an Einstein biography written by Reichinstein. In spite of his promise not to publish this manuscript, Reichinstein later did so [19]. Together with Rudolf Goldschmidt Einstein received a patent (Deutsches Reichspatent Nr. 590783) for an apparatus for people hard of hearing [20].

Out of many well-known men Einstein met in the Physical Society Emil J. Gumbel may be mentioned. He was proposed as a member of the society by Wilhelm Westphal (24. 6. 1921). Gumbel spoke on 17. 2. 1922 (chairman H. Rubens) about "Statistical considerations on the measurements of H. Rubens on the radiation law". Gumbel also reported for the "Physikalische Berichte" from 1924, when he was in Heidelberg. He is well known for his work on the theory of probabilities and statistics, and, as well, for his political commitment, which caused disciplinary actions, reported in the journal „Die Menschenrechte". Einstein was prepared to defend Gumbel. Gumbel's address in the list of members of the society was (1925): Moscow, Marx-Engels-Institute.

The last list of members of the society was published in 1937. There one finds the following names among others: V. Fock, J. Franck, Ph. Frank, G. Herzberg, P. Kapitza, R. Ladenburg, C. Lanczos, I. Langmuir, O. Stern, E. Wigner.

Albert Einstein delivered his last lecture in the German Physical Society on 17. 7. 1931 in the Harnack House in Berlin-Dahlem, at a joint session with the German Society for Technical Physics on the occasion of the death of Albert Abraham Michelson (19. 12. 1852—9. 5. 1931): "Commemorative address on Albert A. Michelson" [21].

Michelson was a member of the German Physical Society and had reported in the session of 16. 6. 1911 "On the construction and the use of bending lattice". Einstein himself met Michelson the last time on 15. 1. 1931 at the California Institute of Technology [22].

In sum, it is clear that the activities of Albert Einstein in the German Physical Society have required a considerable part of his working power during the hard years of the First World War. His work and his place in this society would merit more research and appreciation.

Einstein — school and education [23]

On these subjects Albert Einstein expressed his opinion frankly in more than 50 articles, and essays, speeches, addresses, and messages, and in his famous autobiographical sketch [24]. These ideas of Einstein had deserved more attention in the

Einstein centenary. Owing to his own shameful school experiences under the authoritarian German state, with methods of drilling and doctrinal procedures in school-life and his pleasant impressions of the Swiss Canton-school in Aarau Einstein had given up his original idea of studying engineering.

At the age of $15^1/_2$ Einstein did not pass the examination for entering the ETH (Federal Institute of Technology, or Polytechnic) in Zürich. Such an examination was no exception, but rather the rule for students coming to the ETH. Einstein attended the Canton-school in Aarau for only one year to obtain the school leaving certificate (Matura). Then he started his studies — one year younger than his classmates — in section VIII: Teachers for mathematics and physics.

In 1912/13 Einstein himself trained teachers in this section VIII on the basis of official printed programmes [25]. The director of this section was Einstein's friend and colleague during their studies: The mathematician Marcel Grossmann (1878—1936).

Albert Einstein was a born teacher who found a lot of joy in teaching. His methods were perhaps somewhat unorthodox and unconventional; this he noticed especially in his time as a private and temporary teacher. His pleasure in teaching was also shown in the circle of his friends (e.g. the olympia academy), to his assistants and his coworkers, and occasionally to fellow-travellers too.

Einstein's remarkable educational influences went beyond the scope of schools and had been effective for the whole of human society. His care was always directed towards the potentialities of the younger generation. He gave numerous lessons, although he was not obliged to do so. Einstein himself stood firmly for teaching foreign students, namely Jews from Eastern Europe at the former Friedrich-Wilhelms University in Berlin. Together with the professor of medicine, L. Landau, he had written a letter (19. 2. 1920) to Helfritz, Minister of Science, Arts and Education, asking for courses in physics, mathematics, medicine, botany, history and oriental languages. Einstein wanted these courses of instruction to be recognized as "officially permitted" by the state, because they should be valid in foreign countries [26].

This intention was supported benevolently by the Minister of Education as Einstein mentioned. For "about 200 foreign young people from Russia, Poland, Bulgaria, Romania, and Lithuania, mostly Eastern Jews" the courses were held. Albert Einstein lectured on "Introduction to theoretical physics", and professor James Franck on "Experimental physics". The lectures in mathematics were given by Prof. Issai Schur (infinitesimal calculus), and Privat-Dozent Dr. Rademacher (geometry). These courses of Einstein's supplement the catalogue of his known teaching activities discovered up to now [27]. This supplementation was found after publication of the work for the documentary volumes „Albert Einstein in Berlin 1913—1933" [28].

On these courses, for which no records exist in the archives of the Humboldt University in Berlin, Einstein himself reported in a newspaper "How I became a Zionist" [29]. This article is not listed in any bibliography and represents a further supplement to my addenda [30]. There is another source where Einstein mentioned the organization of lectures by Jewish and non-Jewish colleagues for eastern Jews: Banesh Hoffmann cited in Tel Aviv (GR 7) a part of a speech held by Einstein in England (1921) [31] after having published the article in the Jüdische Rundschau.

Albert Einstein himself felt obliged as a Jew to support the Zionist movement, particularly with the foundation of the Hebrew University of Jerusalem. But at no time did Einstein support chauvinistic aims and methods. In the article mentioned above [29] he wrote: "Often antisemitism is a question of political calculus". He also pointed to the attitude of science in England towards his theory of relativity: "While in general in Germany the judgement of my theory depends on the political position of the newspapers, the attitude of English scientists has demonstrated that their sense of objectivity is not to be blurred by political points of view."

Einstein's understanding of Zionism and his support for the Zionist movement have developed into an increasingly essential element in his life. Undoubtedly it is an important task to comprehend this inescapable side in its circumstances. Moreover it remains to be investigated if there are differences between the definitions of Zionism in Einstein's publications and in present printings, that is, if a change of terms possibly took place. Einstein himself professed, on the one hand, to be in the Zionist movement, which he supported within certain limits. On the other hand, however, Einstein's attitude was characterized as if it followed a PLO-book [32]: "Outstanding Jewish thinkers (including Einstein, Cohen, Rosenwald and Magnes) recognized the racial restrictiveness, the narrow-minded, chauvinistic, isolationist, and supernationalistic elements of Zionism and warned of their consequences. Their criticism is extremely valuable for us, because not all of them were anti-Zionists: Rosenwald really is an outspoken anti-Zionist, but Magnes was Zionist and Einstein and Cohen were more non-Zionists than anti-Zionists".

Einstein's feelings of belonging to the Jewish people first awoke when he was in Berlin. So he wrote to the „Zentralverein deutscher Staatsbürger jüdischen Glaubens" (Central Union of German Citizens of the Jewish Faith): "I am not a German citizen ... I am a Jew and I enjoy belonging to the Jewish people, although I do not think about it in terms of the elect" [33]. In this source not mentioned in a bibliography until now, one can obviously recognize a change in application of the term "race", too.

Einstein's public activities were decisively determined by his remarkable interest in social and educational problems. He attached great significance to the educational factor in social development. Often he lamented over the passivity and cowardice of academicians, "Denkmenschen" (intellectuals) as he called them. He strongly believed in power of reason and thought that the voice of a rational man cannot die away without being heard. With that and with his lifelong commitment to peaceful and humanistic aims, the teacher and educator Albert Einstein is still effective for the future.

Einstein — cultural and intellectual life

Posterity regards the extraordinary creative life of Albert Einstein, so rich in ideas, with great admiration. His excellent scientific reputation was already established in the history of science years before his arrival in Berlin (April 1914). Here in Berlin Albert Einstein increasingly became an immense symbol in public, mainly in the intellectual, life. In comparison to the presentation of his purely scientific works and results his non-scientific activities have been relatively ignored.

Einstein's nature was like that of an artist or prophet. His love of the arts was ethically accented and art made him happy. Occasionally he entered a neighbouring field from motives of art. Einstein — himself a physicist — embodied certain traits and qualities of an artist in the highest sense. The first bust of Einstein was created by Kurt Harald Isenstein (1923) and set up in the Astrophysical Institute (Einstein-tower) [34] at Potsdam in 1929.

Einstein's great love of music and his own violin-playing are well known. His „Lina" — as he called his violin — accompanied him on many travels. Less known is the fact that Einstein himself wrote by hand an opinion of a violin constructed by the violin-manufacturer Erich Kielow (Potsdam). This piece seems to be the last thing Einstein wrote in Caputh on 6. 12. 32 — his luggage was ready for transport, as Mrs. Margarete Kielow reports, to whom Einstein in person gave his manuscript about the violin in his summer-house [35].

Here is the original German text and in translation: „Ich hatte heute Gelegenheit, eine von Herrn Erich Kielow hergestellte Geige zu probieren und mit einer anderen vortrefflichen modernen Geige zu vergleichen. Die Geige des Herrn Kielow ist zweifellos eine der schönsten Geigen, die ich in der Hand gehabt habe. Sie spricht leicht an, hat einen großen runden und ausgeglichenen Ton. Es ist fraglos, daß solche Kunst des Geigenbaues Förderung verdient; das Stadium muß doch endlich überwunden werden, in dem man denkt, daß eine vortreffliche Geige *alt* sein müsse". — (Today I had the opportunity to test a violin constructed by Mr. Erich Kielow and to compare it with another excellent modern one. Without any doubt the violin of Mr. Kielow is one of the most beautiful violins I have ever held in my hands. It responds to a gentle touch, and has a large round and balanced sound. Beyond all question, such an art of violin construction deserves encouragement; the attitude of thinking that an excellent violin has to be *old* must be overcome at last.)

From these handwritten sentences by Einstein one can feel his joy and his art of formulation. Again there appears clearly his practice of calling in question the customs of thinking and traditional conceptions: Why must only old violins be good?

We are grateful to Miss Dukas, Albert Einstein's secretary for 27 years, and to his former coworker Banesh Hoffmann for giving new glimpses in the Einstein Archive. In this book by Dukas and Hoffmann one finds the physicist Einstein's own words on artistic matters. So he said in 1952 about Faraday — with regard to the present time: "This man loved mysterious Nature as a lover loves his distant beloved. In his day there did not yet exist the dull specialization that stares with self-conceit through hornrimmed glasses and destroys poetry ..." [36].

On request, to say something about the close connection between arts and sciences, Einstein summarized his ideas (1921) in an aphorism: "What Artistic and Scientific Experience Have in Common: Where the world ceases to be the scene of our personal hopes and wishes, where we face it as free beings admiring, asking, and observing, there we enter the realm of Art and Science. If what is seen and experienced is portrayed in the language of logic, we are engaged in science. If it is communicated through forms whose connections are not accessible to the conscious mind but are intuitively recognized as valuable, then we are engaged in art. Common to both is the loving devotion to that which transcends personal concerns and volition" [37].

Obviously events of all kinds to do with Albert Einstein were frequently material for the newspapers. Einstein's name shone from advertising pillars in calls and appeals. His name even appeared on a theatre placard, showing his close connection to art and artists. In honour of the deceased great actor Albert Steinrück, the play "The Marquis of Keith" (Frank Wedekind) was staged with many stars on 28. 3. 1929, at 11 p.m., that is, after the end of the usual theatrical performances, in the Schauspielhaus Am Gendarmenmarkt (today: Place of the Academy). The commemorative address was given by Heinrich Mann.

Among the 86 popular actors — many of whom were banished several years later because of their Jewish origin — one finds sonorous names like Max Hansen, Elisabeth Bergner, Fritzi Massary, Fritz Kortner, Eduard von Winterstein, Hans Albers, Ernst Deutsch, Kurt Goetz, Marlene Dietrich, Asta Nielsen, Henny Porten etc.

Albert Einstein is named on this theatre placard in the "Honour Committee" of this extraordinary theatrical performance [38]. Other members in this Committee included Max Reinhardt, Max Liebermann, the Minister of Culture, Prof. Dr. Becker, the Lord Mayor of Berlin, Böß, and the President of the Reichstag, Paul Löbe.

In a less known and easily overlooked document Einstein, the internationalist, referred to a current phenomenon of that time, to nationalism, in a serious scientific journal. His often expressed opinion about famous men and women of science and the arts gave well-founded contributions on questions of the history of science. Once he was requested by his friend Arnold Berliner, editor of „Naturwissenschaften", to write his opinion on Arago's commemorative address about Thomas Young. Einstein's communication concluded with the words: "Not without a certain malicious joy have I noticed, as a child of our generation, while reading Arago's speech, that the men of science already in earlier times were not at all free of the weakness of nationalistic narrow-mindedness; therefore we needn't feel today like exiles from paradise. But I hope that this consolation will not give us full satisfaction" [39].

One receives a lucid and vivid impression of Einstein's personality, of his conceptions, of his method of thinking and of his power of colourful prose expression from his many prefaces and forewords, which are still not fully compiled in bibliographies. Not only has he here articulated critical judgments as hardly any other scholar has — always again worthwhile reading — but also he was not faint-hearted and sparing with self-critical utterances in his self-descriptions and self-assessments.

In his foreword, already written in 1942, to Philipp Frank's book that appeared in 1979 as a reprint of the first edition (1949) in the German language, one recognizes Einstein's critical attitude, writing that biographies "had seldom attracted or captivated" him. About autobiographies he even said that "their origins are usually thanks to the self-love or feelings of negative characters against their fellow-beings". In his typical and constantly refreshing remarks he wrote that "temperament and external circumstances" had lent to his "life something of a coloured exterior — superficially considered". He asked the question whether "one must devote a biography at all to such a life, directed to recognizing and comprehending" [40].

It seems that he himself is giving a hint as to such peculiarities as orderliness and thoroughness: In his review of Elsbach's book "Kant und Einstein" — only recently listed in the bibliography [41] Einstein wrote in his unconventional manner: "Elsbach's book marks itself out by the clear and clean way that its ideas are formed, by its honesty and thoroughness, the last even a little bit too much".

Einstein's intellectual achievement extended beyond physics and, in its versatility and profundity, attached many fields; philosophy also received especially many stimuli through his pioneering knowledge. The philosopher Hans Reichenbach once said: "Einstein's work contains more philosophy than some philosophical systems".

The conception "Copernican Revolution" is firmly established in the history of culture and thought; and in the same manner will be marked „Einstein's Revolution" because he developed cosmology into a branch of physics and gave foundations of the theoretical physics in the 20th century.

Einstein — peace and humanism

The great humanist Albert Einstein recognized the dominance of three great forces playing a negative role in human lives: stupidity, fear and greed. His achievements were directed against these three evils throughout his life. Abhorrence of violence and war induced him to stand for peace, democracy and the progress of mankind, and to raise his voice — always, however, preserving his own independence. He saw clearly that education for peace and humanism is a front-line task in a progressive educational policy.

In November 1914 Albert Einstein was a co-founder of the pacifist organization „Bund Neues Vaterland" (League of the new fatherland), which aimed to fight against jingoism; it was forbidden in 1916. Ninety-one men famous in public and scientific affairs wrote to the former Chancellor, von Bethmann Hollweg, on 27. 7. 1915, among them Albert Einstein. In their statement they affirm "the principle that the annexation or incorporation of politically independent or traditionally independent people is to be rejected". They also rejected means "which would indirectly lead us finally to annexation" [42].

Einstein's pacifism was not pacifism as an end in itself,but in its essence a courageous anti-fascism with many dangers for body and life. This was also true for many of his co-fighters, among them for Emil J. Gumbel, and for the philosopher and mathematician Bertrand Russell, under hard pressure in England. Russell's book "Political ideals" (1922) was translated and introduced by E. J. Gumbel. Albert Einstein wrote in the preface — recently listed in the bibliography [43] with spirit and accuracy: "Not a tottering professor speaking to us, who balances the one thing with another, but one of the decidedly straightforward individuals, standing there independent of the times in which they are more or less accidentally born".

This opinion by the younger Einstein of Russell was true also of himself all his life. The striving for independence of thought and action is so marked for Albert Einstein himself and his oft-made characterizations of outstanding persons as "independent individuals" coincides with his own endeavours.

Einstein's conceptions of morals, ethics and justice and his fearlessness in pursuing humanity, right and truth are still as relevant as at any time before. His human work for peace shows ways and possibilities for the peace of mankind; as his friend and trustee, Prof. Dr. Otto Nathan, gives us insights, lively and true to originals, in Einstein's writings, speeches and letters [44].

Einstein fought with great devotion for total disarmament and for the abolition of war in general. Militarism, armament and war were for Einstein not compatible with the dignity of people. The most prominent physicist of his time fought tirelessly against war and took part in all essential actions for peace.

Just recently a "Call of the Jewish League for Peace" became known after finishing work on the documentary book „Albert Einstein in Berlin, 1913—1933" [45], only briefly mentioned in this volume. Albert Einstein was a member of the presidium of the Jewish League for Peace. Besides Einstein other members were the rabbi Leo Baeck, G. Simon, H. Stern and C. Wassermann as the chairman. In this call of the Jewish League for Peace, also signed by Albert Einstein (1929), we read: "Catholic and Protestant organizations for peace all over the world have invited Jewish groups to get together with them into working associations for peace. With this suggestion, which our community cannot evade comes the inner motive for co-operating in the world historical task of our time... The Jewish League for Peace would like to cooperate with sister-communities in other states and hopes that similar organizations will arise in many countries of culture... Let us unite all our efforts in order to transpose into life everything today which was dreamed of and fought for by our ancestors" [46].

Mahatma Ghandi was highly esteemed and admired by Albert Einstein because of his work for the liberation of India and his demonstration of what sacrifices people are able to suffer if they oppose an apparently unlimited material power. Einstein commended this example and the "way of non-cooperation in Ghandi's sense" e.g. at the time of McCarthy's methods of inquisition (1953) [47]. At this time he composed a preface for a book by Gene Sharp "Ghandi Wields the Weapon of Moral Power" which appeared in 1960; it is still not compiled in bibliographies [48].

Albert Einstein's last years were more than ever filled with sorrow over the future of mankind whose further existence is continually threatened by armament and the danger of war. He was incessantly an admonisher of and conscience for the world because "the thinking of the future must make wars impossible". In his clear articulation, he claimed in 1953 "Mere praise of peace is easy but not effective. What is needed is active participation against war and against everything that leads to it" [49]. Rightly and prophetically he said: "The true problem is lying in the hearts of the people".

Otto Nathan's meritorious book on Einstein's monumental work for peace [50] was wonderfully supplemented by Helen Dukas and Banesh Hoffmann's book "New glimpses from the Einstein Archive" [51]. In it an epigram of Einstein is printed, which was deposited in a capsule in 1936: "Dear Posterity, If you have not become more just, more peaceful, and generally more rational than we are (or were) — why then, the Devil take you.

Having, with all respect, given utterance to this pious wish, I am (or was) Yours, Albert Einstein" [52].

If at any time this capsule is found, the readers will know: This only Albert Einstein could say so strikingly.

Notes

[1] Parts of an unpublished manuscript (1978)

[2] A. Einstein: „Über die Entwicklung unserer Anschauungen über das Wesen und die Konstitution der Strahlung" (Bern, 14. 10. 1909). Verhandlungen der Deutschen Physikalischen Gesellschaft **11** (1909) 482–500; Physikalische Zeitschrift **10** (1909) 817–826.

[3] A. Einstein: Eine neue elektrostatische Methode zur Messung kleiner Elektrizitätsmengen. Phys. Zeitschr. **9** (1908) 216–217.

[4] Conrad und Paul Habicht: Elektrostatischer Potentialmultiplikator nach A. Einstein. Phys. Zeitschr. **11** (1910) 532–535.

[5] A. Einstein, M. Besso: Correspondance 1903–1955. Paris 1972 pp. 46/47. Ed.: P. Speciali.

[6] W. Gerlach: Erinnerungen an Albert Einstein 1908–1930. In: Albert Einstein. Sein Einfluß auf Physik, Philosophie und Politik (Hrsg.: P. C. Aichelburg und R. U. Sexl). Braunschweig/Wiesbaden 1979. I am very indebted to Prof. Gerlach and his widow Dr. Ruth Gerlach for valuable information.

[7] Catalogue of the „Gedächtnisausstellung zum 100. Geburtstag von Albert Einstein, Otto Hahn, Max von Laue und Lise Meitner". Berlin (West) 1979. — I am grateful to Jost Lemmerich, the organizator of this exhibition, for sending me the catalogue.

[8] I am very grateful to Prof. L. Pyenson (Montreal) for helpful hints to my English translations and for giving me a xerox of Einstein's letter to Albert Gockel, published in the article by J. Laub „Albert Einstein and Albert Gockel" Academia Friburgensis **20** (1962) Nr. 1 pp. 30–33.

[9] The famous colloquia — later called Max von Laue colloquia — mostly took place in the same house. Unfortunately these colloquia and Einstein's participations are unexplored by historians of physics. It seems that there do not exist any documents in this matter. It may be in assets of colleagues, but unknown up to now.

[10] See note 5, p. 58.

[11] A. Einstein: Ein einfaches Experiment zum Nachweis der Ampèreschen Molekularströme. Verhandl. Dt. Phys. Ges. **18** (1916) 173–177. Einstein demonstrated this experiment in the session of the Physical Society on 25. 2. 1915. On the 19. 2. 1915 he reported on work carried out with W. J. de Haas: „Experimenteller Nachweis der Ampèreschen Molekularströme"; Verhandl. Dt. Phys. Ges. **17** (1915) 152–170.

[12] All these contributions — together with the reply of Max Planck — were published: E. Warburg, M. v. Laue, A. Sommerfeld und A. Einstein: Ansprachen zu Max Plancks sechzigstem Geburtstag. Müller, Karlsruhe, 1918. — Einstein's contribution „Motive des Forschens" see also Verhandl. Dt. Phys. Ges. **20** (1918) 69.

[13] Verhandl. Dt. Phys. Ges. 1928, p. 15/16.

[14] Carl Seelig: Albert Einstein. Zürich 1960, p. 258–260.

[15] I am very indebted to Prof. Dr. Otto Nathan (Estate of Albert Einstein, New York) for this information.

[16] Edith Einstein: Zur Theorie des Radiometers (thesis, Zürich 1922). Leipzig 1922; published also in Annalen der Physik **69** (1922) 241.

[17] A. Einstein: Theorie der Radiometerkräfte. Zeitschr. Phys. **24** (1924) 1–6.

[18] H. Melcher: Albert Einsteins Patente. Spektrum **9** (1978) Heft 9, 23–26.
H. Melcher: Albert Einsteins technische Erfindungen. Der Neuerer, Heft 5/6 (1979) 202 bis 205.

[19] I am grateful to Prof. O. Nathan for this information too.
[20] See note 18.
[21] A. Einstein: Gedenkworte auf Albert A. Michelson. Zeitschr. f. angewandte Chemie 44 (1931) 685.
[22] Judith Goodstein: Albert Einstein in California. Engineering & Science. California Institute of Technology. May/June 1979, p. 17—19. I am very indebted to Prof. Dr. Max Delbrück (Pasadena) for sending me this journal.
[23] Parts of an unpublished manuscript (1979/80).
[24] A. Einstein: Autobiographisches. In: Albert Einstein als Philosoph und Naturforscher (ed. P. A. Schilpp). Stuttgart 1951.
[25] I wish to express my thank to Dr. Beat Glaus (ETH Zürich, Einstein-Collection) for sending me these programmes and other materials.
[26] A. Einstein und L. Landau: Letter to the Minister für Wissenschaft, Kunst und Volksbildung, Helfritz, (192. 1920). Zentrales Staatsarchiv (GDR), Dienststelle Merseburg, Ministerium des Innern. Akte Rep. 76 Kultusministerium V a Sekt. 1 Tit. VII Nr. 78, Zulassung von Ausländern zum Studium an den Universitäten, Bd. 7 1919—1921.
I am grateful to the director of the Dienststelle Merseburg, Mr. Waldmann, for searching and sending copies.
[27] H. Melcher: Albert Einsteins Lehrveranstaltungen in Berlin. Wissenschaft und Fortschritt 29 (1979) 72—73.
H. Melcher: Albert Einsteins Lehrveranstaltungen an der Philosophischen Fakultät der Friedrich-Wilhelms-Universität Berlin. Potsdamer Forschungen. PH Potsdam, Naturwiss. Reihe, Heft 14, 1979, p. 53—58.
[28] H.-J. Treder and Chr. Kirsten (ed.): Albert Einstein in Berlin 1913—1933, part I and II. Berlin 1979.
[29] A. Einstein: Wie ich Zionist wurde. Jüdische Rundschau, Jg. 26, No. 49 (21. 6. 1921), p. 351—352.
[30] H. Melcher: Addenda zur Einstein-Bibliographie — mit Kommentaren. Potsdamer Forschungen. PH Potsdam, Naturwiss. Reihe, Heft 14, 1979, p. 73—90.
[31] B. Hoffmann: Einstein und der Zionismus. In: Albert Einstein. Sein Einfluß auf Physik, Philosophie und Politik (Hrsg. P. C. Aichelburg und R. U. Sexl). Braunschweig/Wiesbaden 1979; p. 177—184.
[32] Darstellungen zum Palästina-Problem. Palästina-Bücher No. 8. Forschungszentrum — Palästinensische Befreiungsorganisation. Beirut (Libanon) 1968, p. 81.
[33] A. Einstein: Reply upon an invitation of the Central Union of German Citizens of the Jewish Faith. Jüdische Presszentrale III. Jahrgang, No. 11 (24. 9. 1920) p. 5 (Zürich).
[34] Prof. Kurt Harald Isenstein (Copenhagen) wrote me in one of his kind letters that he had created the bust of Einstein already in 1923. I am very grateful to the artist who died only recently.
[35] I am very indebted to Mrs. Margarete Kielow for giving me Einstein's handwritten opinion and giving me a lively report of her visit to the Einsteins in Caputh (6. 12. 1932) which we have tape-recorded. — This opinion was only shortly mentioned in documentary books (see note 28).
[36] H. Dukas and B. Hoffmann: Albert Einstein. The Human Side. New Glimpses from His Archives. Princeton 1979, p. 99; original German text p. 157.
[37] Ibid. p. 37/38; original German text p. 131/132.
[38] B. Engelmann: Deutschland ohne Juden. Eine Bilanz. München 1970, p. 102 ff.; here one finds a copy of the theatre placard.
[39] A. Einstein: Statement on the commemorative address of Arago on Thomas Young. Naturwissenschaften 17 (1929) 363.
[40] A. Einstein: Vorwort of the book of Philipp Frank „Einstein. Sein Leben und seine Zeit". Braunschweig 1979. (This foreword is only contained in this reprint of the edition of the year 1949).
[41] A. Einstein: Review of the book Elsbach, A. C. "Kant und Einstein. Untersuchungen über das Verhältnis der modernen Erkenntnistheorie zur Relativitätstheorie". Berlin und

Leipzig 1924. In: Deutsche Literaturzeitung 1924, 24. Heft, 1685–1692. Bibliography see note 30.
[42] Writing to the Chancellor von Bethmann Hollweg. Zentrales Staatsarchiv (GDR), Dienststelle Merseburg, Ministerium des Innern. Akte Rep. 77, Titel 885, Nr. 4, Bd. 1, f 94, 94 v. I am grateful to the institution in Merseburg for carefully searching and copying.
[43] A. Einstein: Vorwort to the book of Russell, B. „Politische Ideale". Translated from English and introduced by E. J. Gumbel. Deutsche Verlagsgesellschaft für Politik und Geschichte. Berlin 1922.
[44] Nathan, O. und H. Norden (Hrsg.): Einstein über den Frieden. Bern 1975.
[45] See note 28.
[46] The call is published in „Das Jüdische Magazin", No. 1, July 1929. For further detail in this matter I am very indebted to Miss Helen Dukas (Institute for Advanced Study, Princeton), Prof. Dr. M. J. Klein (Yale University) and Dr. Sybil Milton (Leo Baeck Institute, New York).
[47] See literature note 44, p. 545 ff.
[48] A. Einstein: Foreword in Gene Sharp "Ghandi Wields the Weapon of Moral Power". Ahmedabad-14, 1960. — Einstein had written this foreword originally in German in 1953. I am very grateful to Prof. O. Nathan for sending a xerox of the foreword.
[49] Literature see note 44, p. 102.
[50] See note 44.
[51] See note 36.
[52] See note 36, p. 105; German text p. 159.

The publication of the Einstein-texts follows with friendly consent of the Estate of Albert Einstein (New York).

Address: Prof. Dr. H. Melcher
Pädagogische Hochschule „Dr. Theodor Neubauer"
Erfurt/Mühlhausen
DDR - 5010 Erfurt
German Democratic Republic

Physical Sense in Relativity: Max Planck Edits the Annalen der Physik, 1906–1918

L. Pyenson (Montreal)

1. Introduction

The center of gravity for a scientific discipline lies in its periodical press. Disciplinary journals control the quality and direction of research, define acceptable limits to scientific debates, provide a measure for individual achievement, and confer status on trusted advisers. In the case of the strongest journals in a discipline, the editorial hand is firm and the editor's vision, conservative. For these reasons, over the past six or seven generations scientists have often felt called to create new journals that could respond to the needs of one or another emerging specialty. Journals devoted to physical chemistry, colloidal chemistry, applied mathematics, number theory, astrophysics, and even general relativity have in this way come into being.

Since disciplines keep a close rein on publishing, it comes as no surprise that some revolutionary scientific ideas appear in print beyond the reach of disciplinary approbation or censure. Forums for authors with major restructurings to propose have included general scientific journals, the proceedings of learned corporations, popular magazines, and privately printed monographs. At various points in their careers these were the favored publication outlets for Charles Darwin, Oliver Heaviside, Sigmund Freud, Alfred Wegener, and Enrico Fermi. The historian of science is especially interested, then, in situations when an established, disciplinary journal accommodates a theory that fundamentally reorients scientific discourse. In such cases journal editors shape how the innovation is received. A sensitive editor can insure that an otherwise unsettling theory is quickly absorbed into the fabric of existing knowledge [1].

In this essay I consider how during the years 1906—1918 the editor of the most prestigious physics journal in the world evaluated incoming manuscripts treating Albert Einstein's theories of relativity. The journal was the *Annalen der Physik,* the principal publishing outlet for Einstein's own early scientific work. The editor was Max Planck, professor at the University of Berlin and Germany's most distinguished theoretical physicist. In writing to his coeditor Wilhelm Wien, Planck expressed opinions about the work of a large number of his scientist contemporaries who sought to contribute to Einstein's formulations. Planck emerges from this correspondence as a cautious, conservative physicist vitally interested in extending the "revolutionary" theories of relativity [2]. Planck sought to have the theories of relativity accepted because they resolved longstanding problems in classical physics. The principle of relativity, Planck noted in 1910, "removes from the [previously existing] physical world picture the nonessential components brought in only by the circumstance of human perception and habits, and so it purges physics of anthropomorphic impurities originating in the individual nature of the physicist" [3]. By winnowing manuscripts submitted to the *Annalen* and by encouraging work that seemed promising, Planck hoped to soften the "hard struggles" that he knew the theories would have to face [4].

Because this story turns on the character of one scientific journal, I begin by tracing the development of the *Annalen der Physik* from its foundation in the eighteenth century to the beginning of the twentieth. After considering the circumstances surrounding Planck's and Wien's editorial charge, I elaborate on Planck's vision of theoretical physics, especially the theories of relativity, as this vision is revealed in his editorial judgments. A central feature of Planck's approach to relativity involved mistrust of mathematical formalism. He believed that the laws of physics transcended the language, mathematics, in which they were expressed. For Planck in the period around 1910, the formal elegance of a physical proposition mattered less than the extent to which it could be used to treat related problems. "The measure of the worth of a new physical hypothesis," he wrote, "lies not in its vivid expression but in its ability to perform well." In his view performance was related to experimental verification: "All physical questions are decided not by aesthetic points of view but by experiments" [5].

2. The Annalen

Beginning with the end of the nineteenth century, the *Annalen der Physik* has traced its ancestry to the efforts of Friedrich Albrecht Carl Gren, a late eighteenth-century professor of physics and chemistry at the University of Halle. Like many of his physical scientist contemporaries, Gren rose through the early stages of a pharmacy career and received a medical doctorate before turning all his interests to physical sciences. Inspired by the chemical journal of his teacher Lorenz von Crell, in 1790 thirty-year-old Professor Gren brought into being the *Journal der Physik*, a periodical publication for "mathematical and chemical branches of natural science." Within four years the journal began a new series, the *Neues Journal der Physik*, again under Gren's watchful eye. Gren used both journals as vehicles to promote his views in favor of the phlogiston theory of chemical combustion [6].

Upon Gren's death in 1798 his editorial responsibilities fell to Ludwig Wilhelm Gilbert. The same as Gren, Gilbert as a child lost his father. Gren's mother sent him to study at the progressive Philanthropinum, a school in Dessau where the influential pedagogue Johann Bernhard Basedow lectured. Gilbert went on to hear physical sciences at the University of Halle and came under the spell of Gren, nine years his senior. Docent and *ausserordentlicher Professor* in 1795, Gilbert took over direction of Gren's *Journal* upon his mentor's death in 1798. Scientific editing for Gilbert was by no means a single-minded life's focus. During his early years as editor he wrote a three-volume travel guide for Germany. That he was eager to begin editing three years before he received Gren's chair suggests that he received either direct compensation or scholarly renown for his labors. He called his journal the *Annalen der Physik*

[7]. During Gilbert's stewardship, the *Annalen* appeared irregularly and published eclectically.

Gilbert's successor was the distinguished bibliographer Johann Christian Poggendorff. Like Gren apprenticed to an apothecary and like both Gren and Gilbert having as a youth lost his father, young Poggendorff arrived in Berlin to study chemistry. He soon made a mark as a talented electrical experimentalist. Having thought of running a scientific journal, upon Gilbert's death in 1824 twenty-six-year-old Poggendorff presented himself to the publisher, the firm of Johannes Ambrosius Barth, as the next man in charge. His candidacy accepted and his services presumably rewarded by the press, Poggendorff changed the journal's title to the *Annalen der Physik und Chemie*. Only six years after assuming his charge did Poggendorff receive a professorial title from Prussia. After a decade of editing, Poggendorff was awarded an honorary *Dr. phil.* and came to be employed as associate professor of physics at the University of Berlin. With Poggendorff's genius for organization, the journal issued 160 volumes in fifty-two years, most of the articles conforming to the editor's empiricist inclinations [8].

By the time that Poggendorff died in 1877, the physical sciences in Germany had been transformed, and the role of a scientific editor had come to require a new kind of talent, support, and organization. Poggendorff's successor, Gustav Heinrich Wiedemann, belonged to the first generation of physicists, in the modern sense of the word. Son of a Berlin merchant who died when Wiedemann was a boy, the future physicist passed through *Gymnasium* in Berlin. Introduced to physical science by an uncle, Wiedemann then went on to receive in 1847 a doctorate in physics at the University of Berlin. His physics education came in the private laboratory of his experimentalist and empiricist professor, Gustav Magnus, who discouraged his students from pursuing mathematical physics. As a result, Wiedemann studied the works of Siméon-Denis Poisson on his own. Privatdocent at Berlin in 1850, he married a daughter of the chemist Eilhard Mitscherlich's the next year. After twenty years of professorial appointments in physics at Basle, Brunswick, and Karlsruhe, in 1871 he obtained at the University of Leipzig the first German chair in physical chemistry. When Poggendorff died in 1877, the publisher of the *Annalen* approached Wiedemann to become editor [9].

Under Wiedemann's direction the journal emerged as the leading forum for original physics publications in a country that was soon to lead the world in this discipline. The transformation occured because Wiedemann encouraged original contributions and increased the number of theoretical papers. At the same time, the change related to a new administrative arrangement. With Wiedemann's ascension the Berlin Physical Society undertook to contribute to the costs of publication, and it delegated Germany's most distinguished physicist, Hermann von Helmholtz, as its factotum in the editorial office. Beginning in 1877 the title page of the *Annalen* specified this organization, Helmholtz's name appearing in type smaller than that used for the editor, Wiedemann. Helmholtz's death in 1893 resulted in the aging Wiedemann's asking his own physicist son Eilhard to become coeditor. Finding a replacement for the overseer Helmholtz took several years. The new representative of the Berlin Physical Society (from 1898 the *German* Physical Society) appeared on the title page in 1895. He was Max Planck, recently appointed professor in theoretical physics at the University of Berlin.

Gustav Wiedemann died in 1899. The coeditor son, an otherwise undistinguished physicist at the University of Erlangen, did not feel up to the task of carrying on in the absence of his father and mentor. By this time it should have been evident to German physicists that the *Annalen* had grown beyond the point where one physicist could edit it. Such a realization did not surface in either Berlin or Leipzig; the press and no doubt the Physical Society sought out the most promising, established young physicist to continue Wiedemann's work, still under Planck's watchful eye. Perhaps to prevent regional jealousies from emerging, the journal fell into the hands of Paul Drude, recently appointed professor of theoretical physics at the University of Leipzig. In an attempt to distribute responsibility for the journal over a broader segment of the German physical community, Drude had behind him a *Kuratorium*, or council, of five professors of physics: experimentalists Friedrich Kohlrausch, Georg Quincke, Emil Warburg, and Wilhelm Conrad Röntgen, and theoretician Planck. The title of the journal changed to emphasize its status as an organ of physicists: it became, once more, simply the *Annalen der Physik*.

Drude was a natural choice as editor. Son of a physician, in 1887 he received a doctorate for a dissertation on theoretical crystallography directed by physicist Woldemar Voigt at the University of Göttingen [10]. Drude worked as Voigt's assistant until 1894, when he was called to become associate professor of physics at the University of Leipzig. In 1900 he went as successor to renovate and direct the moribund physics institute at the University of Giessen. His institute there was a small one, attracting fewer doctoral students in physics than nearly any other German university, and his budget was commensurate with the institute's low popularity. The position carried few administrative responsibilities [11]. By the time that he went to Giessen, Drude had accumulated a remarkably long and varied list of publications. He was at his finest when he interpreted and extended Maxwell's electrodynamics, as elaborated by Heinrich Hertz. Drude belonged to a tradition exemplary in the work of Helmholtz and Hertz, where theoreticians were also expected to be at home with experimental physics. The marriage of theory and experiment in Drude's published work was more harmonious than that found in the research of any of his distinguished young contemporaries, including Philipp Lenard, Wilhelm Wien, and Emil Wiechert. Having by around 1900 published scores of papers, monographs and textbooks, Drude was seen as the inheritor of Helmholtz's and Hertz's mantle. It was entirely natural that he have been called to direct the journal that had published most of his work. To edit a voluminous and prestigious review, direct a small institute and continue to produce first-rate research was a difficult task, even for someone with Drude's talents. All physicists in Germany looked to Drude's rising star.

When in 1905 Emil Warburg resigned from the University of Berlin to become the third president of the Imperial Institute of Physics and Technology, his position, the most prized chair of physics in Germany, went to Drude. Over the preceding ten years Warburg had set a breathtaking record as institute director. He issued about eight doctoral dissertations a year, more than any other German professor of physics

His students published more than students at any other physics institute, some 220 publications between 1895 and 1905. By the middle 1920s around one-third of his students from this period held professorships [12]. Going from physics at Giessen to physics at Berlin would have implied great changes in one's style of research, teaching, and administration, even without the additional burden of editing the *Annalen*. Drude did not make the transition. His personal research and writing slowed. He was overwhelmed.

Drude called out in anguish to his friend Wilhelm Wien at Würzburg, asking that Wien change places with him. A country boy who liked living in an uncongested city, Wien blanched at the thought of directing the Berlin institute, for it, constructed on piles driven into the banks of the Spree and with a high tension electrical cable running underneath, was entirely unsuited for delicate physical measurements. Finding no honorable way out, and not communicating his desperation to colleagues at Berlin, Drude committed suicide. It was only one year after he had arrived in the imperial city. The shock rippled through the world of physics. Weeks after Drude shot himself, Max Levin, a post-doctoral student at Göttingen, wrote about the event to Ernest Rutherford, professor at McGill University in Montreal. He understated that Drude "was somewhat overworked, but a satisfactory explanation has not been found" [13].

3. Plank and Wien take charge

It was under these circumstances, then, that Max Planck stepped forward to become editor of the *Annalen*. To share editorial responsibilities he asked Wilhelm Wien, then professor of physics at the University of Würzburg in the south German state of Bavaria. As was the case with Drude, Wien's appointment provided visible evidence that the Berlin Physical Society sought to represent all German physicists, those in and beyond Prussia. Wien is best known today for work in synthesizing experimental research which led to the quantum theory of radiation, but his activity spanned all of physics. He was an early and vocal supporter of the electromagnetic view of nature and an elaborator of the electron theory. Like Planck, Wien became an immediate supporter of Einstein's special theory of relativity. Different from Planck, in the 1920s Wien appeared in the company of the anti-Semitic, anti-relativity physicists Johannes Stark and Philipp Lenard [14].

In a letter to Wien written in 1906, Planck proposed how the new *Annalen* would be managed. The Berlin theoretical physicist wanted both his and Wien's name to appear side by side on the journal's title page, as was the case for the *Zeitschrift für physikalische Chemie*, edited by Wilhelm Ostwald and Jacobus Henricus van't Hoff. Wien would handle the day-to-day matters associated with the journal, although when a manuscript was to be rejected or revised, Planck had to be consulted. Beginning his association with the most prestigious physics journal in the world, Wien asked Planck about the proportion of manuscripts that had been rejected by Drude's editorial hand. Planck could not supply precise figures, but he estimated that only 5% to 10% of submissions had been returned to authors [15]. With hindsight this remarkable statistic helps to explain the appearance in the *Annalen* of a consistent quantity of dull, unoriginal, and insignificant articles. Although under Planck's and Wien's direction the rejection rate seems to have risen (an educated guess would place it at around 15% or 20% in 1914), no clearer indication than this can be provided of the extent to which physics publication in Wilhelmian Germany was available to almost any determined and flexible author.

Planck closed his letter with the hope that Wien would soon receive word from Friedrich Althoff, the powerful civil servant who supervised professorial appointments at all institutions of higher learning in Prussia [16]. As he clarified two days later, Planck had in mind that Wien comes to Berlin as Drude's successor, a position that Wien found attractive. Wien, who did receive but declined the call, worried about the enormous responsibilities entailed by such a position, in conjunction with editing the *Annalen*. Planck assured Wien that Drude had had the possibility of diminishing his work load, but that he had made no move to do so [17]. Though we lack the letters that Wien wrote to Planck, it is clear that the younger man at Würzburg consistently deferred to his senior colleague. To the extent that he wanted to be involved with it, Planck controlled the *Annalen*. Holding the journal firmly in rein, Max Planck shepherded colleagues toward the new physics of the twentieth century.

The public Planck projects an image of a distant, superior sage. Even in the few instances when he reflected on his life, as in his scientific autobiography, personal remarks were with rare exceptions foregone. In corresponding with his coeditor Wien, recipient of the 1911 Nobel prize in physics (Planck would receive it only seven years later), Planck allowed a bit of his private side to show. Planck enjoyed writing letters. As his own research slowed because of advancing age and administrative commitments, his scientific correspondence swelled, and he found the circumstance "enormously stimulating and invigorating" [18]. Business documents, Planck's letters to his near-peer telegraphed succinct judgments about manuscripts by authors knocking on the door of the *Annalen*. To Wien Planck expressed himself in a way that he could never allow in a publication. "Completely without value ... nothing new ... contradictions" are evaluations that issue from Planck's pen. One is struck by how these comments on the substance of manuscripts are distinguished in Planck's letters from his evaluation of personality and character. Planck divorced the business of physics in the *Annalen* from personal questions, insofar as he was able. Even to his coeditor of some twenty years, Wilhelm Wien, Planck never entirely warmed up. The two always addressed each other as „Sie".

From this brief description of Planck's temperament it follows that he would have gone to great lengths to keep caustic polemics from appearing in his journal. A controversy in the *Annalen* was not a pleasant affair. A regular contributor, Einstein, wrote in 1910 to his young colleague, Paul Hertz, that he wanted to speak with Hertz about the latter's recent publication (probably on the mechanical foundation of thermodynamics) rather than address a reply for publication. "A quarrel in the *Annalen*", Einstein wrote to Hertz, "is not a laughing matter" [19]. We can see the extent to which Planck strove to avoid controversy on a personal level from Planck's advice to Wien, in 1906, that the *Annalen* reject a manuscript of Carl Wilhelm Max Koppe's on the concept of relative motion and the Foucault pendulum. The manuscript represented an attempt by Koppe, a fifty-three-year-old professor at the Andreas Realgymnasium in Berlin and a long-time contributor to the proceedings of the Berlin mathematical society and to the *Zeitschrift für physikalischen und chemischen Unterricht*, to join a debate in the pages of the *Physikalische Zeitschrift* over an article on the same subject by Polish physicist Alfred Denizot [20]. In Planck's view, Koppe's article would be "superfluous" for *Annalen*. At the same time, Planck feared that if published it could give rise to a fearsome controversy. Denizot had previously had a manuscript rejected by the *Annalen*, and Planck no doubt felt that Denizot would have reason to claim persecution at the hands of the journal [21]. Apprised of Planck's feeling that his manuscript contributed nothing new, Koppe replied that he really wanted to have the paper appear in the *Annalen*. Planck wrote to Wien that a way out would be to accept Koppe's paper on the condition that he rewrite it to exclude mention of Denizot's work [22]. Koppe dropped the matter and sent a version of his article to the *Physikalische Zeitschrift*, where it appeared immediately [23].

The odyssey of a manuscript submitted in 1906 on the principle of relativity and electromagnetism, written by Alfred Heinrich Bucherer, indicates how Planck and Wien processed articles through their journal. Bucherer was an unusual German physicist who, after having studied at several universities in the United States, returned at an advanced age to take a doctorate at the University of Berlin. In a series of short communications and in an elementary textbook, Bucherer sought around 1905 to contribute to the exciting and mathematically elaborate discussion on the electron theory. He worked apparently oblivious of recent, sophisticated publications by Karl Schwarzschild, Paul Hertz, and Arnold Sommerfeld [24]. Planck carefully scrutinized Bucherer's submission. It was a mess. The coeditor of the *Annalen* found that according to Bucherer's interpretation of the principle of relativity, a moving current of air would impart its velocity to a light wave, a result in contradiction with Fizeau's classic experiment. Even worse, Bucherer did not seem aware that Maxwell's equations held for any uniformly moving system. Because Bucherer was a privatdocent and had worked on Kaufmann's experiments, Planck was in favor of leniency. He urged a revision rather than outright rejection. Generosity was especially indicated, Planck noted to Wien, because Drude had previously rejected a paper of Bucherer's on thermoelectric fluids [25]. Bucherer, however, refused to make changes in his paper, and he asked for a collective opinion by the curators of the *Annalen* [26]. The matter passed to Planck and Wien's "overseers". They opted to support the editors, and wrote to Bucherer about their decision. Bucherer replied that he would not entertain a compromise, as Planck had advocated. Planck hoped, with this response, that the matter would die, and that Bucherer would not in the future come to the *Annalen* [27]. Bucherer's thoughts went no farther than a preliminary paper published previously in the *Physikalische Zeitschrift* [28]. A number of years later Bucherer claimed to have verified the Lorentz theory of electrons by measurements of Becquerel rays, to the uninformed delight of mathematician Hermann Minkowski and the unbridled skepticism of experimentalist Alfred Bestelmeyer [29].

Historian Stanley Goldberg has shown how elaboration and verification of special relativity remained foremost in Planck's mind during the years before 1910 [30]. Planck emerged as one of the very first physicists to extend Einstein's work, and between 1905 and 1914 he was the principal or supplementary adviser for more than a dozen doctoral dissertations that were based at least in part on Einstein's special theory of relativity [31]. When Einstein's paper appeared in print, Planck had already expressed interest in the limits of the mechanical explanation of electrical phenomena, for in 1905 he was principal adviser of a dissertation by Hans Witte on precisely this subject [32]. In the wake of Einstein's work, Planck encouraged his student Kurd von Mosengeil to pursue a theoretical investigation of relativistic thermodynamics. In 1906 Planck saw Mosengeil's dissertation through press and revised the text for the *Annalen* after the premature death of his student [33].

Planck by no means limited himself to theoretical studies. One of his charges was Erich Hupka, officially working under the direction of experimentalist Heinrich Rubens. From the acknowledgment in his dissertation it is clear that theoretician Max Planck provided much guidance for Hupka's attempts in 1908 and 1909 to measure the change in electron mass with electron velocity. Other experimentalists had attempted to obtain such precise measurements, but none of the results were unambiguous. Hupka wanted to provide a definitive decision between the predictions of Max Abraham's theory of the rigid electron and the predictions of the Lorentz-Einstein theory (which Hupka called, along with many others of the day, not „Relativitätstheorie" but „Relativtheorie" — "relative theory" instead of "relativity theory"). Hupka worked with cathode-rays, then established to consist of electrons moving at velocities approaching that of light. Negatively charged cathode-rays were deflected by a magnetic field, the amount of deflection depending only on the apparent electron mass. The young physicist could establish the kinetic energy of electrons emitted from a cathode in a vacuum tube, and he could calculate, for a given magnetic field strength, deflections of the cathode-rays according to Max Abraham's theory and the relativity theory. His observed deflections fitted the latter [34].

When he published his dissertation in monograph form and as an article in the prestigious *Annalen der Physik,* Hupka found himself at the center of a sharp controversy with Wilhelm Heil, who had just finished a dissertation under Planck's direction which critically examined Walter Kaufmann's measurements of the change in electron mass with electron velocity for beta rays [35]. Taking into account the reliability of the data, Heil concluded that experimental evidence did not provide a conclusive decision among the three competing electron theories: those of Bucherer, Abraham, and the „Relativtheorie". Planck had the two doctoral candidates working in ignorance of each other. According to a letter that Planck wrote to Wilhelm Wien, at the time that Heil finished the young researcher did not know of Hupka's work [36]. Heil wrote a sharp critique of Hupka's dissertation and sent it to Planck for publication in the *Annalen.* Planck naturally felt that Heil's subject was "very important", but he urged Heil to moderate his language. Planck informed Hupka about the impending publication. He worked with both researchers to eliminate personal remarks from their position papers [37]. Their public discussions resolved little.

4. The gatekeepers

Planck's attitude toward mathematics and especially how he distinguished mathematical formalism from physical reasoning is clearly revealed in his editorial correspondence with Wien. The Berlin theoretician, of course, was the very model of a physicist brahmin. A university professor like his father, Planck grew up in an atmosphere redolent with the responsibilities and prerogatives of professorial station. His interests turned almost exclusively toward abstract learning, many steps removed from direct contact with the world of practical activity. Planck's research reflects in physics the widespread desideratum of nineteenth-century German, neoclassically-inspired learning, where one was expected to elaborate on the world in "general" terms. Culture was to be *allgemein*, general, rather than *fachlich* or *realistisch*, specialized or practical. Generality implied a primary emphasis on linguistic skills, in philology and in natural sciences [38].

Though a master of mathematical methods, Planck passionately sought to express the fundamental laws of the universe in words. From the fundamental laws, he believed, could be constructed what he and others called a worldpicture of physical reality. It would be as a vast landscape, not unlike those projected by nineteenth-century, German, neoclassical artists, wherein all parts of physics stood in harmony with each other. When words failed him and he held only mathematical formulas — as seems to have been the case in 1900 upon his first formulation of the quantum theory of radiation — he was unable to draw unambiguous conclusions [39]. Planck had little patience with mathematically pretentious glosses on the principle of relativity. Into such a category fell about half of the relativity manuscripts that passed across his desk. His thoughts on several submissions are especially illuminating in this regard.

In 1908 Emil Kohl, associate professor of physics at the University of Vienna, submitted a two-part manuscript that developed a new theory of electrodynamics and critiqued the Michelson experiment. Kohl assumed that electricity was a continuous fluid distributed throughout space. He came to the same results as those obtained by Lorentz, Planck noted, but only after having made special hypotheses about the ether. Planck urged Wien to ask Kohl to limit his observations to the Michelson experiment. The outcome was as he requested [40]. In rejecting a later manuscript of Kohl's that set out a theory of electrons, Planck emphasized that among all Kohl's many equations he had not found "a single one in which a new relationship between measurable quantities is provided" [41]. Kohl is the physicist who in 1911 was edged out by Einstein for a chair at the German university in Prague [42]. In a similar class was a manuscript of Anton Weber's on special relativity. In Planck's view it did not have "enough physical results to be accepted by the *Annalen*." It would make only "ballast" for the journal [43]. Weber, a professor of physics and mathematics at the Royal Bavarian Lyzeum in Dillingen, was only able to make his thoughts public in a note published in the *Physikalische Zeitschrift* [44].

Planck considered as "entirely worthless" two long manuscripts submitted in 1911 by Emil Arnold Budde on the Klinkerfues and Michelson experiments to detect motion relative to the ether. Sixty-nine-year-old Budde directed the Charlottenburg factory of the firm Siemens & Halske; he had published extensively in the *Annalen* and had in 1888 directed the abstracting journal *Fortschritte der Physik*. Budde wrote in the style of an engineer. Planck found that Budde was completely ignorant of the literature and that his attempted critiques of the two experiments were embarrassingly bad. His work contained "no original thought that is not already found in the scientific literature, and done better there" [45]. Both of Budde's papers, rejected by Planck, appeared in the *Physikalische Zeitschrift*. In his paper on the Michelson experiment, Budde criticized Max von Laue's textbook of 1911 on the special theory of relativity. Laue replied to Budde's accusations with devastating effect [46].

On a manuscript of F. Grünbaum's which was ultimately rejected, Planck commented in 1911 that it was "correct, but it includes nothing really new and its physical interest is only very indirect." The paper duplicated a lecture that applied mathematician Hans von Mangoldt had published in the *Zeitschrift* of the German Engineer's Association and reprinted in the *Physikalische Zeitschrift* [47]. Planck was not clear if or how Grünbaum used Mangoldt's work, and whether he supplied anything more than mathematical formulas to Mangoldt's physical content. Grünbaum's article appeared shortly thereafter in the *Physikalische Zeitschrift* [48].

To judge from its contents the biweekly *Physikalische Zeitschrift*, controlled by Göttingen physicists and in this period edited variously by Emil Bose, Friedrich Krüger, Hans Busch, Max Born, and Heinrich Theodor Simon, was often desperately short for copy. Publishing both notes and long-winded analyses, the journal became a dumping ground for work rejected by Planck and Wien. Even so, some manuscripts declined by the *Annalen* did not find their way into the more catholic journal, presumably because the treatments were obviously derivative or out of fashion. One such case was a long manuscript elaborating Vilhelm Bjerknes's hydrodynamical analogue for electromagnetism, submitted in 1912 by a certain H. Rudolph. In 1910 Rudolph had published a small book purporting to unite the principle of relativity, Planck's quantum of radiation, and gravitation in a mechanical picture of the world [49]. Planck would not have Rudolph's elaboration of this theory. Bjerknes's mechanical theory had appeared in the *Annalen*, along with a rejoinder by Hans Witte, but "direct and definitive rejection" was Planck's advice for Rudolph's manuscript, a text that failed to distinguish between force and pressure and one that remained confused about the physical meaning of differential quotients [50].

The preceding papers were all written by unimportant authors whose work was far from original. Not all submissions were so easily weighed. In 1910 Planck reluctantly acceded to a manuscript by Waldemar Sergius von Ignatowsky on the notion of a relativistic rigid body. Ignatowsky in fact met with Planck and told Planck that Wien was not happy with his manuscript. Planck commented on Ignatowsky's confusion over Einstein's notion of signal velocity, but in the end decided to accept Ignatowsky's paper [51]. Planck had to handle Ignatowsky with care, because Ignatowsky and Eugen Jahnke — both aging privatdocents at the Berlin Institute of Technology — had proposed to create a journal specializing in theoretical physics, a competitor for many articles that would otherwise be sent to the *Annalen*. Planck approached the project, which did not bear fruit, with circumspection. To Wien he confided that it might be "quite a good idea" to take some theoretical work out of the *Annalen*, but he dreaded the consequent emergence of "a sharp division between theoretical and experimental research." He believed that theory had always to be grounded in experimental reality [52].

It appears from Planck's correspondence with Wien that the most perplexing submissions on relativity were those invoking complicated mathematical machinery to elaborate formal, working hypotheses. Planck especially believed that the *Annalen* had to adopt a clear policy with respect to submissions dealing with the principle of relativity. Manuscripts that focused on the formulation of definitions — as was the case in the recent spate of literature on the relativistic rigid body — had to be referred to mathematical journals or to the more accommodating journal *Physikalische Zeitschrift* [53]. Planck urged that a 1913 manuscript by the twenty-three-year-old Polish physicist Felix Joachim de Wiśniewski be declined. "The author defines every last thing in a formal way and assumes that behind it all these definitions have a physical meaning. But nothing new comes from it." Wiśniewski's gravitational theory might have had some strong points, but in Planck's view there were "too few solid, deciding factors for a completely informed gravitational theory." At this time Planck believed that even Einstein's theory was not necessarily in the right direction, and it would have to be tested during the upcoming solar eclipse of 1914 [54]. In two previous papers published in the *Annalen*, Wiśniewski had begun to elaborate a new gravitational theory, but Planck decided that the journal did not have to continue to support Wiśniewski's tedious and pedestrian mathematical speculations [55]. A second communication on the quantum theory, submitted by Wiśniewski in 1914, also received definitive rejection by Planck [56].

For Planck, mathematical exposition had to be clear as well as relevant to physical concerns. In 1913 he accepted one short paper from Jun Ishiwara [57], a Japanese theoretical physicist who had studied extensively in Europe, but later that year Planck convinced Wien to reject another of Ishiwara's papers on electrodynamics. The second treatment contained serious mathematical infelicities, such as defining one quantity without further comment as a "Quasisinnevektor". In the expositions, as in other publications of Ishiwara's, the author was not always clear and the text would have to be rewritten completely. Planck did not want to hurt Ishiwara's feelings. He suggested to Wien that in rejecting the manuscript one could say that it was not publishable in the present form. In any event, Ishiwara had already published the result in a Japanese journal. In all probability Ishiwara sent the rejected manuscript to the *Physikalische Zeitschrift*, where it appeared in 1914 [58].

The above extracts tend to present Planck as a stern gatekeeper. In reality he encouraged work that he thought promising, even if it did not issue from the pens of his students. He followed Walther Ritz's emission theory of radiation with great interest, even though he did not believe in it [59]. In 1908 Planck advised Wien to accept a paper that the young Viennese physicist Philipp Frank had submitted, where Frank showed how the Lorentz transformation could reduce to a Galilean transformation and applied the principle of relativity to Hertz's equations for moving bodies. Planck was in favor of the paper even though he remained unclear about the distinction between Einsteinian and Hertzian relativity as elaborated by Frank

[60]. He felt that his journal was fortunate to have Breslau *Oberlehrer* Ferencz Jüttner's "quite interesting" research on kinetic molecular theory and relativity. Planck urged Wien not to cut the manuscript of one of Jüttner's two papers [61]. The senior editor in Berlin urged that a paper on gravitational theory by Finnish physicist Gunnar Nordström be accepted even though it did not offer "fundamental" insights. Nordström was a talented man who had previously not appeared in the *Annalen*, and Planck wanted to encourage Nordström's work. He was especially glad that in Nordström's paper the foreigner retained the constancy of the velocity of light, a principle that Einstein and Abraham had recently dropped. The *Annalen* had to be hospitable, in Planck's view, to promising first communications [62].

After the covariant field equations of general relativity emerged late in 1915, Planck found the *Annalen* besieged by authors wanting to contribute to the topic. In March 1916 Einstein sent the *Annalen* a long article setting out the definitive form of general relativity [63], but many others who knocked at Planck's door with texts elaborating the theory were far removed from centers of power and prestige in the discipline. Einstein gave the wide-ranging engineering professor at Berlin Hans Reissner "many explanations and criticisms," and so helped him complete a paper on the self-gravitation of an electrical field [64]. The young Viennese theoretician Friedrich Kottler elaborated in 1916 the principle of equivalence in a short paper printed without much editorial deliberation [65]. Both Reissner and Kottler had previously published on relativity and gravitation. Planck also argued in 1916 that two manuscripts by the Norwegian physicist Thorstein Gunnar Wereide be accepted, even though as a foreigner and, according to Planck, an "autodidact," Wereide proceeded in an unorthodox manner and wrote with many spelling mistakes. Wereide had published, the previous year, a monograph in English which summarized many of his ideas [66]. One of the papers that Wereide sent Planck, on energy exchange between ether and matter, borrowed from Niels Bohr's atomic theory. Planck urged that it be published because in such a new field standards were different from those in older fields. The manuscript had been rejected by the *Physikalische Zeitschrift*, Planck noted, and that journal's poor judgment was a boon for the *Annalen* [67].

Among the many manuscripts sent to the journal, which elaborated general relativity came one from Königsberg *Oberlehrer* Ernst Reichenbächer, according to Planck a "basically cultured theoretician," who attacked the general problem of the connection between electricity and gravitation. Reichenbächer limited his study to a two-dimensional field which he then expanded to the fourdimensional world of Hermann Minkowski. Planck was sympathetic with Reichenbächer's approach, but he was not overly sanguine about its future. "The value of such a theory," Planck felt, lay in "what it finally delivers." The payoff, in Planck's view, lay in "simplicity and intuitiveness [*Einfachheit und Anschaulichkeit*] and above all in whether it has such characteristic consequences that can be tested by experiment." Reichenbächer's theory failed on both counts. Planck found especially perplexing a law of Reichenbächer's where the radius of curvature of a negative electron was enormously larger than the electron radius. In general Planck felt that the theory was not terribly new if one was familiar with the theory of conformal mappings in two planes. In Planck's view the manuscript was not yet ready for publication. The first, mathematical part had to be clarified; the second, physical part had to deal with the theories of Gustav Mie, David Hilbert, and Einstein [68]. Reichenbächer's manuscript went back to him. Three months later a revision arrived on Planck's desk. Planck was uncertain to which of Einstein's papers Reichenbächer appealed. He urged that Reichenbächer speak with Einstein and so resolve their differences. The meeting was amicable [69]. Reichenbächer's paper appeared in 1917 as the first attempt at a unified field theory in the wake of Einstein's covariant field equations.

Near the end of the war, the problem of mathematical expositions came to weigh heavily on Planck's shoulders. When in 1917 Hermann Weyl sent the *Annalen* his first attempt at a unified field theory, Planck wrote to Wien that Weyl stood at the very "height of research of his time." Although he observed that Weyl did not cite all the literature and mentioned nothing about experimental verification of the theory, Planck noted with approval that Weyl based his work firmly on Einstein's "general gravitational theory." Studies like Weyl's were, in Planck's view, of clear value, but a larger problem remained. Weyl's paper depended heavily on mathematical machinery from non-Euclidean geometry, and Planck would have preferred to see more weight attached to physical reasoning and discussion. He did not want to decide in general the extent to which studies like Weyl's belonged in the *Annalen*, although he offered that possibly "non-Euclidean geometry, as such, separated from physical tasks, will be treated better in mathematical journals as has been the case until now" [70].

5. Plank the editor

Scientific editing calls many kinds of people. In pursuing riches some pander to public tastes. Others seek a special outlet for a particular kind of wisdom or a learned corporation. All scientific editors purport to instruct; their enterprise is an educational one. So it is with Max Planck, an exemplary teacher. In addition to helping to produce a many doctoral dissertations on the theories of relativity, he corresponded with Wien about the submissions of as many as a score of additional authors writing on relativity. The *Annalen der Physik* was controlled by robust and young researchers at the height of their abilities. No different from other people, physicists mature in their positions, but this circumstance is no reason to burden editors with the image of exhausted thinkers. The editors of the *Annalen der Physik* continued educational and scientific activity at the same time that they processed the work of their colleagues. The most valuable commodity at their disposal, time, went to imposing their prejudices on the visible and permanent residue of their discipline-learned publications.

Although Planck was wary in approaching mathematical formalism, he remained in awe of talented mathematician colleagues. He wrote to Wien in 1912 that Hilbert's radiation theory was quite interesting from the point of view of formalism and general applicability, but that it brought no new physical understanding. "For all that," Planck offered, "it is to be welcomed when the mathematicians begin to be interested in physical problems" [71]. Planck was not alone in his opinion about the role of mathematics in physics. His coeditor colleague Wilhelm Wien approached mathematics in a similar way. Wien wrote to David Hilbert in 1909 about his sadness at the death of mathematician Hermann Minkowski, whose last papers on relativity theory, "in which he went entirely into physical points," were of great interest [72]. Like Wien, Albert Einstein would have been sensitive to Planck's strictures. Tensor analysis came to him as a method of last resort. "You have absolutely no idea," Einstein wrote to physicist Paul Hertz in 1916, "what I went through as a mathematical ignoramus until I arrived in this harbor" [73]. The widely travelled young Paul Ehrenfest, later Einstein's close friend, shared this view of mathematics. Ehrenfest wrote to Paul Hertz around 1906 about how he had taught himself higher mathematics, and so his education had many holes: "Often quite elementary mathematical methods are essentially unknown to me." In another letter to Hertz from this period, Ehrenfest emphasized that a surprising majority of the talented physicists and mathematicians whom he had met considered "mathematics a 'veritable devil' — naturally a man-eating one." Ehrenfest added: "I calculate with this fleeting intimation instinctively, and I have the conviction that you must have quite often [experienced] the same sentiment" [74].

The attitude of these physicists toward the role of mathematics in formulating physical laws stands in sharp contrast to that of younger theoretical physicists in the period after the first world war. "Physical sense" was for the younger men increasingly seen to be of less importance than the requirement that a theory be clothed in elegant mathematics. Writing to Wolfgang Pauli about Pauli's long essay on the theories of relativity, the septuagenarian mathematician Felix Klein reported the belief of his mathematician colleague David Hilbert, "that one could explain the essence of nature by mere mathematical reflection" [75]. Hilbert's attitude came to permeate physics in the 1920s. Unfamiliar mathematical expressions replaced classical physical notions, and theorists like Werner Heisenberg, Wolfgang Pauli, and Paul Adrian Maurice Dirac imputed new physical meaning to sophisticated mathematical expressions. Niels Bohr convinced physicists to accept a new, indeterminist epistemology that could accommodate the success of formal methods in quantum mechanics. Those sharing an older vision, however, hesitated to accept the new point of view and, with a few exceptions, refrained from contributing to the structure of the new world picture.

In view of his persistent belief in many features of the late nineteenth-century "physical world picture," Planck appears as a sympathetic figure striding across two epochs. He consistently pointed the way to the new physics of relativity and quanta; in this regard his pedagogical and epistolary activities were as valuable as his original scientific communications. At the same time he resisted abandoning beliefs about physical reasoning and the use of mathematical tools which he had acquired when in the nineteenth century he wrestled with the foundations of thermodynamics. Especially in his role as editor of the *Annalen der Physik*, Planck acted as Moses for twentieth-century physicists. He guided and disciplined his colleagues through nearly twenty years of bewildering revelations, but he never touched the soil of the promised land.

Notes

[1] On the role of editors in scientific publishing see Susan Sheets-Pyenson, Low Scientific Culture in London and Paris, 1820—1875 (diss., University of Pennsylvania, 1976), University Microfilms International no. 77—10, 216, and her subsequent studies: "War and Peace in Natural History Publishing: The Naturalist's Library, 1833—1843", Isis 72 (1981), 50—72; "From the North to Red Lion Court: The Creation and Early Years of the Annals of Natural History", Archives of Natural History, 10 (1981), 221—249; "Darwin's Data: His Reading of Natural History Journals, 1837—1842", Journal of the History of Biology, 14 (1981), 231—248; "A Measure of Success: The Publication of Natural History Journals in Early Victorian Britain", Publishing History, 9 (1981), 21—36.

[2] Max Planck, ,,Die Stellung der neueren Physik zur mechanischen Naturanschauung", Physikalische Zeitschrift, 11 (1910), 922—932, on p. 928. Stanley Goldberg's lucid treatment of Planck's views is never far from discussion here. Goldberg, "Max Planck's Philosophy of Nature and His Elaboration of the Special Theory of Relativity", Historical Studies on the Physical Sciences, 7 (1976), 125—160.

[3] Planck, "Stellung" (note 2), p. 931.

[4] Ibid., p. 930.

[5] Ibid., pp. 929, 931.

[6] Karl Hufbauer, "Gren, Friedrich Albrecht Carl", Dictionary of Scientific Biography, 5 (New York 1972), 531—533. Karl Hufbauer, "The Formation of the German Chemical Community, 1720—1795" (Berkeley, 1982), pp. 120—137.

[7] Ludwig Choulant, ,,Versuch über Ludwig Wilhelm Gilbert's Leben und Wirken", Ann. Phys. Chem., 76 (1826), 453—471.

[8] W. Baretin, ,,Johann Christian Poggendorff", Ann. Phys. Chem., 160 (1877), v-xxiv; Friedrich Klemm, ,,Poggendorff, Johann Christian", Dictionary of Scientific Biography, 11 (New York 1975), 49—51.

[9] Hans-Günther Körber, ,,Wiedemann, Gustav Heinrich", Dictionary of Scientific Biography, 14 (New York 1976), 529—531.

[10] Stanley Goldberg, "Drude, Paul Karl Ludwig", Dictionary of Scientific Biography, 4 (New York 1971), 189—193.

[11] Information on doctoral students from Lewis Pyenson and Douglas Skopp, "Educating Physicists in Germany circa 1900", Social Studies of Science, 7 (1977), 329—366, on p. 350; information on budgets from Paul Forman, John L. Heilbron, and Spencer Weart, Physics circa 1900: Personnel, Funding, and Productivity of the Academic Establishments [volume 5 of Historical Studies in the Physical Sciences] (Princeton: Princeton University Press, 1975), p. 61.

[12] Hans Ramser, ,,Warburg, Emil Gabriel", Dictionary of Scientific Biography 14 (New York 1976), 170—172; Pyenson and Skopp, op. cit. (note 11), p. 355.

[13] Wilhelm Wien, "Aus dem Leben und Wirken eines Physikers" (Leipzig, 1390), pp. 24—26. In the notes to his remarkable novel "Night Thoughts of a Classical Physicist" (Cambridge, Mass., 1982), p. 200, Russell McCormmach emphasizes that "Drude never asked for help, and so Planck and others concluded he didn't neer or wish it". Max Levin to Ernest Rutherford, 25 July 1906. Rutherford Correspondence. Rare Book Department, McGill University, Montreal.

[14] Hans Kangro, ,,Wien, Wilhelm Carl Werner Otto Fritz Franz", Dictionary of Scientific Biography, 14 (New York 1976), 337—342.

[15] Planck to Wien, 28 July 1906. Handschriftenabteilung, Staatsbibliothek Preussischer Kulturbesitz, Berlin (West) [henceforth SPK Berlin].

[16] Ibid.

[17] Planck to Wien, 30 July 1906. SPK Berlin. See Wilhelm Wien, ,,Aus dem Leben und Wirken eines Physikers" (Leipzig: J. A. Barth, 1930), pp. 24—26. This letter of Planck's is noted in McCormmach, "Night Thoughts" (note 13), p. 200.

[18] Max Planck, Scientific Autobiography and Other Papers, trans. Frank Gaynor (New York: Philosophical Library, 1949), pp. 49–50.
[19] „... eine Streiterei in den Annalen ist kein hübscher Spass". Albert Einstein to Paul Hertz, postcard dated 15 August 1910. Collection of Rudolf H. Hertz, Roslyn Heights, New York, and available on microfilm in the Einstein Archives, Firestone Library, Princeton, New Jersey.
[20] Alfred Denizot, „Zur Theorie der relativen Bewegung und des Foucaultschen Pendelversuches", Phys. ZS., **6** (1905), 342–345; L. Tesař, „Die Theorie der relativen Bewegung und ihrer Anwendung auf Bewegungen auf der Erdoberfläche", ibid., 556–559; Denizot's reply, ibid., 559 and 677–679; P. Rudzki, „Theorie der relativen Bewegung", ibid., 679–680.
[21] Planck to Wien, 28 July 1906. SPK Berlin.
[22] Planck to Wien, 12 October 1906. SPK Berlin.
[23] Max Koppe, „Zum Foucaultschen Pendel", Phys. ZS., **7** (1906), 604–608, 665–666.
[24] Alfred H. Bucherer, „Das Feld eines rotierenden Elektrons", Phys. ZS., **6** (1905), 225 bis 227; „Die Rotation eines Elektrons mit Volumladung", ibid., 269–270; „Das deformierte Elektron und die Theorie des Elektromagnetismus", ibid., 833–834; Mathematische Einführung in die Elektronentheorie (Leipzig: Teubner, 1904). Bucherer's work is considered in Goldberg, "Planck's" (note 2), 129–132. The research front in electron theory circa 1905 is studied at length in Lewis Pyenson, "Physics in the Shadow of Mathematics: The Göttingen Electron-Theory Seminar of 1905", Archive for History of Exact Sciences, **21** (1979), 55–89.
[25] Planck to Wien, 29 November 1906. SPK Berlin.
[26] Planck to Wien, 21 December 1906. SPK Berlin.
[27] Planck to Wien, 26 January 1907. SPK Berlin.
[28] Alfred H. Bucherer, „Ein Versuch, den Elektromagnetismus auf Grund der Relativbewegung darzustellen", Phys. ZS., **7** (1906), 533–557.
[29] Alfred H. Bucherer, „Messungen an Becquerelstrahlen. Die experimentelle Bestätigung der Lorentz-Einsteinschen Theorie", Phys. ZS., **9** (1908), 755–762, including the debate following Bucherer's presentation at the 1908 meeting of the Naturforscher in Cologne.
[30] Goldberg, "Planck's" (note 2).
[31] Between 1906 and 1914 Planck was principal director for the following dissertations which in whole or part related to relativity and the electron theory: Hans Witte, „Über den gegenwärtigen Stand der Frage nach einer mechanischen Erklärung der elektrischen Erscheinungen", Abschn. 1: „Begriff, Grundlagen, Einleitung" (1905); Kurd von Mosengeil, „Theorie der stationären Strahlung in einem gleichförmig bewegten Hohlraum" (1906); Walter Meissner, „Zur Theorie des Strahlungsdruckes" (1907); Wilhelm Heil, „Zur Theorie der Kaufmannschen Versuche über die elektromagnetische Ablenkung der β-Strahlen" (1909).
[32] Witte, „Erklärung" (note 31).
[33] Goldberg, "Planck's" (note 2), 133–137.
[34] Erich Hupka, Die träge Masse bewegter Elektronen (Dr. phil. diss., University of Berlin, 1909).
[35] Wilhelm Heil, Zur Theorie der Kaufmannschen Versuche über die elektromagnetische Ablenkung der β-Strahlung (Dr. phil. diss., University of Berlin, 1909).
[36] Planck to Wien, 30 November 1909. SPK Berlin.
[37] Planck to Wien, 6 July 1910. SPK Berlin. Erich Hupka, „Beitrag zur Kenntnis der trägen Masse bewegter Elektronen", Ann. Phys., **31** (1910), 169–204; Wilhelm Heil, „Diskussion der Versuche über die träge Masse bewegter Elektronen", ibid., 519–546.
[38] Lewis Pyenson, "Mathematics, Education, and the Göttingen Approach to Physical Reality, 1890–1914", Europa (Montreal), **2** (1979), 91–127, pp. 101–102; Mathematics and Neohumanism in Nineteenth- and Early Twentieth-Century Germany, to appear as a volume in the Memoirs series of the American Philosophical Society, Philadelphia.
[39] Thomas S. Kuhn, Black-Body Theory and the Quantum Discontinuity, 1894–1914 (Oxford: Clarendon, 1978). In general, Kuhn's book does not speak to the issues raised here.

[40] Planck to Wien, 26 November 1908. SPK Berlin. Emil Kohl, ,,Ueber den Michelsonschen Versuch“, Ann. Phys., 28 (1909), 259–307.
[41] Planck to Wien, 30 November 1909. SPK Berlin.
[42] József Illy, ,,Albert Einstein in Prague“, Isis, 70 (1979), 76–84.
[43] Planck to Wien, 7 November 1910. SPK Berlin.
[44] Anton Weber, ,,Konvektions- und Röntgenstrom in der Relativitätstheorie“, Phys. ZS., 11 (1910), 134.
[45] Planck to Wien, 30 May 1911. SPK Berlin.
[46] Emil Budde, ,,Das Dopplersche Prinzip für bewegte Spiegel und ein Versuch von Klinkerfues“, Phys. ZS., 12 (1911), 725–729; ,,Zur Theorie des Michelsonschen Versuches“, ibid., 979–991, including discussion of Budde's paper at the 1911 Naturforscherversammlung in Karlsruhe.
[47] Hans von Mangoldt, ,,Längen- und Zeitmessung in der Relativitätstheorie“, Phys. ZS., 11 (1910), 937–944.
[48] Planck to Wien, 9 February 1911. SPK Berlin.
F. Grünbaum, ,,Ueber einige ideelle Versuche zum Relativitätsprinzip“, Phys. ZS., 12 (1911), 500–509.
[49] H. Rudolph, Die mechanische Erklärung der Naturerscheinungen, insbesondere der Relativbewegung, des Planck'schen Wirkungselements und der Gravitation (Coblenz: W. Groos, 1910).
[50] Planck to Wien, 29 June 1912. SPK Berlin. Hans Witte, ,,Besonderes und Allgemeines zur Weltätherfrage: eine Erwiderung auf einen Artikel von Herrn V. Bjerknes“, Ann. Phys., 32 (1910), 382–410.
[51] Woldemar von Ignatowsky, ,,Der starre Körper und das Relativitätsprinzip“, Ann. Phys., 33 (1910), 607–630. Planck to Wien, 13 July 1910. SPK Berlin.
[52] Planck to Wien, 13 June 1910. SPK Berlin.
[53] Planck to Wien, 9 February 1911. SPK Berlin.
[54] Planck to Wien, 29 June 1913. SPK Berlin.
[55] Felix de Wiśniewski, ,,Zur Minkowskischen Mechanik", Ann. Phys., 40 (1913), 387–390, 668–676.
[56] Planck to Wien, 14 June 1914. SPK Berlin.
[57] Planck to Wien, 31 July 1913. SPK Berlin. Jun Ishiwara, ,,Ueber das Prinzip der kleinsten Wirkung in der Elektrodynamik bewegter ponderabler Körper“, Ann. Phys., 42 (1913), 986–1000.
[58] Planck to Wien, 14 December 1913. SPK Berlin. Jun Ishiwara, ,,Grundlagen einer relativistischen elektromagnetischen Gravitationstheorie“, Phys. ZS., 14 (1914), 294–298, 506–510.
[59] Planck to Wien, 19 January 1909. SPK Berlin.
[60] Planck to Wien, 9 October 1908. SPK Berlin. Philipp Frank, ,,Das Relativitätsprinzip der Mechanik und die Gleichungen für die elektromagnetischen Vorgänge in bewegten Körpern“, Ann. Phys., 27 (1908), 897–902.
[61] Planck to Wien, 9 February 1911. SPK Berlin. Ferencz Jüttner, ,,Das Maxwell'sche Gesetz der Geschwindigkeitsverteilung in der Relativtheorie“, Ann. Phys., 34 (1911), 856–882; ,,Die Dynamik eines bewegten Gases in der Relativtheorie“, ibid., 35 (1911), 145–161.
[62] Planck to Wien, 28 January 1913. SPK Berlin. Gunnar Nordström, ,,Träge und schwere Masse in der Relativitätstheorie“, Ann. Phys., 40 (1913), 856–878.
[63] Planck to Wien, 6 March 1916. SPK Berlin. Albert Einstein, ,,Die Grundlage der allgemeinen Relativitätstheorie“, Ann. Phys., 49 (1916), 769–822.
[64] Hans Reissner, ,,Ueber die Eigengravitation des elektrischen Feldes nach der Einsteinschen Theorie“, Ann. Phys. 50 (1916), 106–120. Reissner credits Einstein for assisting him.
[65] Friedrich Kottler, ,,Ueber Einsteins Aequivalenzhypothese und die Gravitation“, Ann. Phys., 45 (1916), 955–972.
[66] Thorstein Wereide, Statistical Theory of Energy and Matter (Kristiania [Oslo]: Gyldendal, 1915).

[67] Planck to Wien, 1 March 1916. SPK Berlin. Thorstein Wereide, „Die statistisch-mechanische Grundlage der allgemeinen Quantentheorie“, Ann. Phys., **49** (1916), 966–975; „Der Energieaustausch zwischen Materie und Aether“, ibid., 976–1000.
[68] Planck to Wien, 25 August 1916. SPK Berlin.
[69] Planck to Wien, 12 November 1916. SPK Berlin. Ernst Reichenbächer, „Grundzüge zu einer Theorie der Elektrizität und der Gravitation“, Ann. Phys., **52** (1917), 134–173.
[70] Planck to Wien, 10 August 1917. SPK Berlin. Hermann Weyl, „Zur Gravitationstheorie“, Ann. Phys., **54** (1918), 117–145.
[71] Planck to Wien, 4 October 1912. SPK Berlin.
[72] Wilhelm Wien to David Hilbert, 15 April 1910. Nachlass Hilbert, Niedersächsische Staats- und Universitätsbibliothek, Göttingen, Federal Republic of Germany.
[73] Einstein to Paul Hertz, 22 August 1916. Collection of Rudolf H. Hertz, Roslyn Heights, New York, and available on microfilm in the Einstein Archives, Firestone Library, Princeton, New Jersey.
[74] Paul Ehrenfest to Paul Hertz, two undated letters, Collection of Rudolf H. Hertz, Roslyn Heights, New York.
[75] Felix Klein to Wolfgang Pauli, 8 May 1921. A. Hermann, K. von Meyenn, V. F. Weisskopf, eds., Wolfgang Pauli. Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg u. a. Band I: 1919–1929 (New York: Springer, 1979), p. 31.

Address: Prof. Dr. L. Pyenson
Institut d'histoire et de sociopolitique des sciences
Université de Montréal
Montréal H3C 3J7
Canada

APPENDIX

Short version of the program

SUNDAY (13. 7. 80)

9.00:	Registration of participants opens *Place:* University Tower
20.00:	E. Schmutzer unveils a bust of Einstein by the sculptor Jo Jastram *Place:* Foyer of Lecture theatre 1/House 1, Max-Wien-Platz 1 (for invited guests only)
Evening:	Informal get-together in the University Tower

MONDAY (14. 7. 80)

Scientific Program

Opening Session
(with accompanying persons)

10.00–10.30:	Welcoming address by the Rector of Friedrich Schiller University, Jena Prof. Dr. sc. med. Dr. med. h. c. F. Bolck Welcoming address in the name of IUPAP by the President of the National Physics Committee of the GDR Prof. Dr. sc. nat. J. Auth (Humboldt University, Berlin) Opening of the Conference by the President of the International Society for General Relativity and Gravitation Prof. Dr. rer. nat. Dr. rer. nat. h. c. P. G. Bergmann (Syracuse University, New York)

Opening Lectures

10.30–11.15:	J. A. Wheeler (Austin): Einstein's second century
11.15–12.00:	E. Schmutzer (Jena): Prospects for Relativistic Physics *Place:* Volkshaus

Plenary lectures

Chairman: W. B. Bonnor (London)

14.00—14.45: D. Kramer and H. Stephani (Jena):
Exact solutions of Einstein's field equations (read by D. Kramer)

14.50—15.30: I. Robinson (Dallas):
Complex methods in General Relativity

15.35—16.10: R. A. d'Inverno (Southampton):
Computer methods in General Relativity

16.45—17.30: J. E. Marsden (Berkeley):
Initial value problems and dynamics of gravitational fields

17.35—18.20: J. Ehlers (Munich):
Equation of motion, gravitational radiation and corresponding approximative methods
Place: Lecture theatre 1/House 1

20.00: Meeting of the International Committee on General Relativity and Gravitation
Place: Café in the University Tower

Non-scientific Program

15.00—17.00: Guided tour of the University Main Building
Guide: H. Hölzel (Jena)
(free of charge)

20.30—22.00: Physics experiments for entertainment
Demonstrator: H.-D. Jähnig (Jena)
Place: Lecture theatre 1/House 1
(free of charge)

TUESDAY (15. 7. 80)

Scientific Program

Plenary lectures

Chairman: I. Novikov (Moscow)

9.00—9.30: R. S. Ward (Dublin):
Present state of the twistor program

9.35—10.20: H.-J. Seifert (Hamburg):
Basic theory of black holes, singularities and topology

10.55—11.40: R. Sunyaev (Moscow):
Compact objects under astrophysical aspects
(read by Ya. B. Zeldovich)

11.45—12.30: Ya. B. Zeldovich (Moscow):
Theoretical and empirical situation in cosmology
Place: Lecture theatre 1/House 1

Chairman: D. Ivanenko (Moscow)

14.00—14.45: R. D. Reasenberg and I. I. Shapiro (Cambridge/MA)
Terrestrial and planetary experiments, time and length standards
(read by R. D. Reasenberg):

14.50—15.35: G. W. Gibbons (Cambridge/UK):
Quantization about classical background metrics

15.40—16.25: P. van Nieuwenhuizen (Stony Brook):
Gauge quantum theories of gravitation

17.00—17.40: S. Ferrara (Frascati):
Supergravity
Place: Lecture theatre 1/House 1

Non-scientific Program

10.00—12.00: Guided tour of the places and objects of interest in the town and within the University (Goethehaus, Schillerhaus, Collegienhof, Botanical Garden, Hanfried-Monument)
Guide: H. Hölzel (Jena)
(free of charge)

14.00—17.30: Trip to Cospeda to the museum and terrain of the Battle of Jena and Auerstedt (1806)
Guide: S. Schmidt (Jena)

15.00—20.00: Exhibition of instruments manufactured by Kombinat VEB Carl Zeiss (primarily for scientists)
Place: Opposite the main entrance of the University Tower (Verhandlungszentrum)
(free of charge)
(Repeat without guide on 16. 7. and 17. 7. from 9.00—15.00)

17.00—18.30: Lecture on "Socialist economic policy — as an example Kombinat VEB Carl Zeiss Jena"
Chairman: W. Matthies

20.30: Concert given by the Jena Philharmonic Orchestra and Singakademie Choir
Conductors: G. Blumhagen and S. Nordmann
Place: Volkshaus

WEDNESDAY (16. 7. 80)

Scientific Program

Discussion groups

9.00—10.30: Exact solutions of Einstein's field equation
Leading moderator: H. Stephani (Jena)
Place: Lecture theatre A1/House 4

9.00—10.30: Terrestrial and planetary experiments, time and length standards
Leading moderator: R. D. Reasenberg (Cambridge/MA)
Place: Lecture theatre 1/House 1

9.00—10.30: Astrophysics of compact objects
Leading moderator: B. Carter (Meudon)
Place: Lecture theatre 4/House 2

11.00—12.30: Equation of motion, gravitational radiation and corresponding approximative methods
Leading moderator: P. Havas (Philadelphia)
Place: Lecture theatre A1/House 4

11.00—12.30: Complex methods in General Relativity
Leading moderator: I. Robinson (Dallas)
Place: Lecture theatre A4/House 4

11.00—12.30: Continuous signal antennae, and Doppler spacecraft ranging for gravity wave detection
Leading moderator: J. Weber (Maryland)
Place: Lecture theatre 1/House 1

Non-scientific Program

10.00—12.00: Guided tour of places and objects of interest in the town and within the University
(repeat of previous day's tour)

10.00—12.30: Excursion to the Karl Schwarzschild Observatory, Tautenburg
(mainly for scientists)
Guide: S. Marx (Tautenburg)

13.30: Visit to Buchenwald Memorial and Weimar

13.45: Excursion to Weimar (only)

20.30: A choice of musical entertainment
(free of charge)

1. Quartet of the Akademische Orchestervereinigung
   Place: Collegienhof, Kollegiengasse
   (in the event of bad weather: Rehearsal Hall of Philharmonie, August-Bebel-Str. 4)

2. Concert with the Jena Philharmonic Madrigal Group
   Conductor: G. Blumhagen (Jena)
   Place: Auditorium of the University Main Building
3. Old Time Memory Jazz Band
   Leader: G. Mlynski (Jena)
   Place: „Rosenkeller"
   Students' Club, Johannisstraße 13

THURSDAY (17. 7. 80)

Scientific Program

Discussion groups

9.00—10.30: Initial value problems and dynamics of gravitational fields
Leading moderator: Y. Choquet-Bruhat (Paris)
Place: Lecture theatre 1/House 1

9.00—10.30: Parametric transducers for gravity wave antennae
Leading moderator: D. Blair (Western Australia)
Place: Lecture theatre 4/House 2

9.00—10.30: Historical aspects of the Theory of Relativity
Leading moderator: H. Melcher (Erfurt)
Place: Lecture theatre A1/House 4

11.00—12.30: Twistors and other approaches to space-time structure
Leading moderator: R. Penrose (Oxford)
Place: Lecture theatre 4/House 2

11.00—12.30: Laser experiments on gravitational waves
Leading moderator: R. W. P. Drever (Glasgow)
Place: Lecture theatre 1/House 1

11.00—12.30: Computer methods in General Relativity
Leading moderator: R. A. d'Inverno (Southampton)
Place: Lecture theatre A1/House 4

14.00—15.30: Quantization about classical background metrics
Leading moderator: P. C. W. Davies (London)
Place: Lecture theatre 1/House 1

14.00—15.30: Relativistic thermodynamics and statistics
Leading moderator: G. Neugebauer (Jena)
Place: Lecture theatre 4/House 2

14.00—15.30: Resonant detectors for gravitational waves (Weber bars)
Leading moderator: D. H. Douglass (Rochester)
Place: Lecture theatre A1/House 4

16.00—17.30: Classical gauge type theories of gravity
Leading moderator: F. Hehl (Cologne)
Place: Lecture theatre 4/House 2

16.00—17.30: Black holes, singularities and topology
Leading moderator: J. Stewart (Cambridge/U.K.)
Place: Lecture theatre 1/House 1

16.00—17.30: Quantum non-demolition detectors
Leading moderators: V. B. Braginsky (Moscow) and C. Caves (Pasadena)
Place: Lecture theatre A 1/House 4

Non-scientific Program

9.00—11.00: Various guided tours (free of charge)

1. Phyletic Museum (museum of human descent)
   Guide: H.-O. Vent (Jena)
2. Museum of Optics
   Guide: F. Rossi (Jena)
3. Karl Liebknecht Memorial
   Guide: H. Hölzel (Jena)
4. Ernst Haeckel House
   Guide: E. Krausse (Jena)
5. Zeiss Planetarium
   Part 1: Through Time and Space in the Planetarium
   Speaker: L. Meier (Jena)
   Part 2: The Heavens of the Relativists
   Concept: J. Dorschner, J. Gürtler, J. Rose (Jena)
6. Bus tour of Neulobeda
   Guides: G. Schulz, City Architect of Jena and M. Berg (Jena)

10.00—12.30: Excursion to the Karl Schwarzschild Observatory, Tautenburg (mainly for scientists)
Guide: S. Marx (Tautenburg)

12.30—15.00: Excursion to the Karl Schwarzschild Observatory, Tautenburg (Group 1, mainly for accompanying persons)
Guide: S. Marx (Tautenburg)

15.00—17.30: Excursion to the Karl Schwarzschild Observatory, Tautenburg (Group 2, mainly for accompanying persons)
Guide: S. Marx (Tautenburg)

12.30—18.00: Excursion to Großkochberg House with a concert in the historic amateur theatre

20.00—1.00: Social evening for all:
„Thüringer Schlachtfest“ (banquet)
Place: Mensa (Refectory)/Philosophenweg 20

FRIDAY (18. 7. 80)

Scientific Program

Plenary lectures

Chairman: F. Bopp (Munich)

9.00—9.40: A. Trautman (Warsaw):
The Einstein-Cartan theory

9.45—10.15: D. R. Brill (College Park) and P.-S. Jang (Syracuse, N.Y.):
Positive energy program
(read by D. Brill)

10.50—11.50: V. B. Braginsky (Moscow) and K. S. Thorne (Pasadena):
Present state of the experiments on gravitational waves

11.55—12.35: L. P. Grishchuk (Moscow):
Experiments on gravitational waves with electromagnetic detectors

Place: Lecture theatre 1/House 1

Discussion groups

14.00—15.30: Supergravity, renormalization program
Leading moderator: P. van Nieuwenhuizen (Stony Brook)
Place: Lecture theatre 1/House 1

14.00—15.30: Cosmology
Leading moderator: I. Novikov (Moscow)
Place: Lecture theatre 4/House 2

14.00—15.30: Alternative classical theories of gravitation; Mach's principle
Leading moderator: N. Rosen (Haifa)
Place: Lecture theatre A1/House 1

16.00: General Meeting of the Society
Place: Lecture theatre 1/House 1

20.30: Meeting of the International Committee on General Relativity and Gravitation
Place: Café in the University Tower

Non-scientific Program

9.00—18.00: Excursion to the Dornburg Houses and to Naumburg, including a visit to the Cathedral and a walk round the Town

9.00—20.00: Excursion to Eisenach (visits to Bach Museum, Luther House, Wartburg Castle) and back to Jena through the Thuringian Forest via Oberhof

SATURDAY (19. 7. 80)

Day of departure

POST CONFERENCE TOURS

Tour I: Jena—Potsdam—Berlin (capital of GDR)

Potsdam:	Visit to Cäcilienhof and the Palace and Park of Sanssouci
Berlin:	Bus tour of the city Visit to the Pergamon museum and the Television Tower
Duration:	Saturday 9.00 until Monday 9.00
Accommodation:	both nights in Berlin

Tour II: Jena—Wernigerode (Harz)—Magdeburg—Berlin (capital of GDR)

Wernigerode:	Walk round the town and visit to the Feudal museum
Berlin:	Bus tour of the city Visit to the Pergamon museum
Duration:	Saturday 9.00 until Monday 9.00
Accommodation:	Magdeburg and Berlin

Tour III: Jena—Meißen—Dresden—Berlin (capital of GDR)

Meißen:	Visit to the pottery demonstration workshop and china display hall
Dresden:	Bus tour of the city Visit to the Zwinger Art Gallery
Berlin:	Bus tour of the city Visit to the Pergamon museum
Duration:	Saturday 9.00 until Monday 14.00
Accommodation:	Dresden and Berlin

Distribution of the participants per country

Algeria	1 + 2	Japan	8 + 2
Argentina	1	Republic of Korea	2
Australia	13 + 2	Mexico	6 + 1
Austria	12 + 2	Netherlands	1
Belgium	4	New Zealand	2
Berlin-West	10	Kenya	1
Brazil	1	Nigeria	1
Bulgaria	5	Norway	4 + 4
Canada	26 + 4	Philippines	1
Chile	4	Poland	22 + 2
China	2	Portugal	1 + 1
CSSR	11 + 1	Rumania	10 + 1
Denmark	1 + 1	Spain	16 + 2
Egypt	1 + 1	Sri Lanka	1
Finland	3	Sweden	11 + 1
France	21 + 4	Switzerland	6 + 3
GDR (DDR)	112 + 7	Taiwan	1
FRG (BRD)	38 + 11	Turkey	2
Greece	3 + 3	United Kingdom	50 + 8
Hong-Kong	1	USA	69 + 10
Hungary	7 + 1	USSR	38 + 5
Iceland	1	Venezuela	1
India	10 + 1	Vietnam	1
Iran	1	Yugoslavia	1
Ireland	5 + 2		
Israel	4		589 + 89
Italy	35 + 9		

The figures after the plus sign indicate the accompanying persons. This list shows that 589 + 89 = 678 persons from 51 countries (Berlin-West and Hong-Kong included) took part in GR 9 as fully admitted participants. Furthermore, 150 persons from the GDR were admitted only for the scientific program. This means the total number of 828 participants.

Grants

Australia	1
Berlin-West	3
Bulgaria	1
Egypt	1
Greece	2
India	18
Ireland	2
Italy	2
Mexico	3
New Zealand	1
Nigeria	1
Poland	4
Rumania	2
Spain	3
Turkey	2
United Kingdom	6
USA	3
USSR	1
Yugoslavia	1
	57

These 57 grants (accomodation, meals, conference fees) have been awarded to all participants who applied for them on time (31/1/80). A few grants were split up in order to satisfy all wishes.

GR 9-poem by N. V. Mitskievich (Moscow)

Ein theoretisch-physikalischer Schiller-Traum

Ach, aus Schwerkrafts tiefsten Gründen,
die das kalte Denken drückt,
könnt' ich doch Effekte finden —
ach, wie fühlt' ich mich beglückt!
Dort erblick' ich Widerscheine,
ewig jung und ewig fein!
Hätt' ich Flügelpaar' alleine,
könnt' ich bald am Ursprung sein.

Harmonien hör' ich klingen,
Töne süßer Theorie;
und die Kongruenzen bringen
mir News Functions schön und früh.
Strenge Lösungen dort erglühen,
winkend zwischen Gleichungsschar'.
Theoretiker sich bemühen
zu verstehen sie ganz und gar.

Ach, wie schön muß sich's ergehen
dort in Newman's Himmelsraum.
Und Lichtkugeln auf den Höhen —
O, wie lab'n sie meinen Traum!
Doch mich wehrt der Killing-Vektor,
der ergrimmt lichtähnlich wird;
Schwarze Löcher scheinen echter,
Singularität verwirrt.

Ein' Mutmaßung fühl' ich schwanken,
aber Formalismus fehlt...
Frisch hinein und ohne Wanken!
Ihre Segel sind beseelt.
Du mußt glauben, du mußt wagen,
denn die Götter leih'n kein Pfand.
Nur Begeisterung kann tragen
dich ins schöne Wahrheitsland.

(Moscow, July 11, 1980)

A Theoretical Physicist's Schiller-Dream

O, out of gravity's deepest grounds
which cold thought keenly presses down
could I but the effects discover –
with what joy would my soul resound!
Forever young, forever new
reflected in the mirror's view.
O, if only wings were mine,
to the source I'd fly in a twinkle of time.

Harmonies do I hear ringing
sweetly melodies of theories singing;
and News Functions early and bright
bring new congruences to my sight.
There wink between equations' masses
exact solutions – a subtle glow passes.
Theoreticians try hard as they should
they understand so well and good.

O it must be quite angelic
there in Newman's heavenly space!
A ball of brightness from on high
nourishes fully my dreaming place.
But now the Killing vector fierce
turned light-like—angrily nears;
Aweful and real the black holes are –
singularity confuses so far.

Assumptions wander through my mind –
but no formalism of that kind!
Dare without a drop of doubt
let fly its sail, o soul, about.
You must have faith and you must dare,
else will the Gods refuse your pawns.
Courage, enthusiasm will carry you there
Sweet Land of Truth – to you be borne.

(Moscow, July 11, 1980)
(translated by Alice Honig
and Anita Ehlers)