Physica **96A** (1979) 70–88 © *North-Holland Publishing Co.*

SCHWINGER AND STATISTICAL PHYSICS: A SPIN-OFF SUCCESS STORY AND SOME CHALLENGING SEQUELS*

PAUL C. MARTIN

Department of Physics and Division of Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

Some of the influences Julian Schwinger has had on condensed matter physics are discussed. The first part rapidly summarizes the language and methods he introduced to describe physical systems exactly and to approximate their properties systematically. The significance of these methods and the ways in which they have been extended are noted. The second part describes how these concepts have been applied to the condensed Bose fluid (i.e., helium 4), a system with rich and varied properties. Some fundamental features of this system are summarized. The third part examines recent advances in our understanding of helium at its critical point in three dimensions and below the critical point in two. A final section describes briefly certain features of chaotic behavior and what is needed to explain them. The problems encountered in the study of turbulence and other chaotic phenomena are compared and contrasted with those arising in other areas of statistical physics. Throughout, the direct and indirect contributions Schwinger has made to condensed matter physics and the contributions condensed matter physics and field theory have made to one another are emphasized.

1. Introduction

During the late 1940's and early 1950's Harvard was the home of a school of physics with a special outlook and a distinctive set of rituals. Somewhat before noon three times each week, the master would arrive in his blue chariot and, in forceful and beautiful lectures, reveal profound truths to his Cantabridgian followers, Harvard and M.I.T. students and faculty. Cast in a language more powerful and general than any of his listeners had ever encountered, these ceremonial gatherings had some sacrificial overtones – interruptions were discouraged and since the sermons usually lasted past the lunch hour, fasting was often required. Following a mid-afternoon break, private audiences with the master were permitted and, in uncertain anticipation, students would gather in long lines to seek counsel.

During this period the religion had its own golden rule – the action principle – and its own cryptic testament – On the Green's Functions of Quantized Fields[1]). Mastery of this paper conferred on followers a high priest status. The testament was couched in terms that could not be questioned, in a language whose elements were the values of real physical observables and their correlations. The language was enlightening, but the lectures were exciting because they were more than metaphysical. Along with structural insights,

* Supported in part by the National Science Foundation through DMR 77-10210.

succinct and implicit self-consistent methods for generating true statements were revealed. To be sure, the techniques were perturbative, but they were sufficiently potent to work when power series in the coupling constant failed because, for example, the coupling was strong enough to produce bound states.

In the dark recesses of the sub-basement of Lyman Laboratory, where theoretical students retired to decipher their tablets, and where the ritual taboo on pagan pictures could be safely ignored, students scribbled drawings that disclosed profound identities between diagrams and sums of diagrams. Few papers have had so large an influence as these papers and the subsequent, less cryptic, version[2]) of part of their content in the series Quantized Theory of Fields, I–VI. Clarifying, justifying, and rephrasing the ideas and the techniques that they contain has occupied many physicists and the results of these activities have often been valuable.

A few years later, in Birmingham and Copenhagen, Cyrano DeDominicis and I turned our hand to the nuclear many-body problem on which work by Keith Brueckner had aroused interest. While we were engaged in this project, Gell Mann and Brueckner were making strides in understanding the quantum electron plasma and Bardeen, Schrieffer, and Cooper were explaining superconductivity. That these three problems had many common features and that a language and techniques akin to those that Schwinger had introduced for relativistic fields should also be developed for equilibrium systems gradually became apparent to both of us. In France, with Claude Bloch, DeDominicis set out to develop a general framework, while at Harvard, upon my return in 1957, I was fortunate enough to enlist Julian's collaboration in the pursuit of this goal.

The paper[3]) Julian and I wrote in 1958 seems to be the only paper of the nearly 200 his bibliography contains that falls in the area of statistical and solid state physics. But it is far from his only contribution to the field. A number of the seventy students whose doctoral research was directed by Julian worked on theses in solid state and plasma physics and several more have gone on to apply tools and modes of thinking he developed in these fields. Thus, although Julian may not realize the degree to which his techniques and their extensions have pervaded the field, I am revealing nothing new to him when I report that field theoretic methods are extremely valuable for studying nonrelativistic many body systems. He and some others among you are likely to be more surprised by the fact that there has also been "spin-off" in the opposite direction, that is, that information about bizarre and unsuspected field theoretic phenomena have emerged from theoretical studies of superfluid helium films, superconductors, and magnetic materials such as $RbMnF_3$, K_2NiF_4, and $LiTbF_4$.

My talk is divided into three parts. The first few minutes are devoted to the language to which I referred. Since it says almost anything, without further specifications, it says almost nothing concrete. If any of you fail to recognize the words, just consider it a mystical incantation recited to exhibit part of the lore Julian generated and why it has permeated almost all of physics.

The second part consists of a case history – what we have learned about interacting Bose fluids, in particular, liquid helium, and how these discoveries have elucidated the rich content of one field theory. I shall talk about discoveries concerning the behavior of such a system at and far below the temperature, T_c, at which it becomes a superfluid.

The third part of the talk continues the discussion of helium, dwelling on aspects of that problem that cannot be even qualitatively explained without adding essentially new methods and ideas to the self-consistent "perturbative" techniques that work elsewhere. Even simpler problems that pose essential difficulties are noted. Common to them all is the absence of a simpler manageable model onto which they can be smoothly mapped.

2. The method

Suppose, for concreteness, we have an interacting boson gas characterized by the Lagrangian density

$$-\mathscr{L} = \frac{\hbar^2}{2m} \nabla\psi^+ \cdot \nabla\psi - \mu\psi^+\psi + \tfrac{1}{2}\lambda\psi^+\psi^+\psi\psi - j^*\psi - j\psi^+ - \tfrac{1}{2}\, \mathrm{i}\hbar(\psi^+\dot{\psi} - \dot{\psi}^+\psi),$$

where $\psi(r)$ and $\psi^+(r')$ satisfy the canonical commutation relations, μ is the chemical potential, j^* a particle source, and λ an interparticle interaction. For comprehensibility, the arguments of the variables have been eliminated; with a sufficiently concise implicit notation in which space time and spin indices of a matrix λ have been suppressed, the formulas describe not only a four point contact potential, but a real non-local two-body interaction. Since explicit treatment of such features leads to equations festooned with indices and obscures the essential ideas, we shall restrict ourselves to the special case of a local four-field interaction when convenient.

With this Lagrangian, the equation of motion is

$$\left(-\mathrm{i}\hbar \frac{\partial}{\partial t} - \frac{\hbar^2\nabla^2}{2m} - \mu\right)\psi + \lambda\psi^+\psi\psi = j$$

from which a functional equation for the action, W, can be generated by noting that with

$$e^W\langle\psi\rangle = \frac{\hbar}{i}\frac{\delta}{\delta j^*}e^W,$$

$$\frac{\hbar}{i}\left(i\hbar\frac{\partial}{\partial t} + \frac{\hbar^2}{2m}\nabla^2 + \mu\right)\frac{\delta}{\delta j^*}e^W + \lambda\left(\frac{\hbar}{i}\right)^3\frac{\delta}{\delta j}\frac{\delta}{\delta j^*}\frac{\delta}{\delta j^*}e^W = j\,e^W.$$

The solution of this equation yields

$$\exp W[jj^*] = \int \mathscr{D}[\alpha\alpha^*]\exp\left(\frac{i}{\hbar}\int \mathscr{L}(\alpha\alpha^*)\,dt\right).$$

From this action a calculational arsenal can be generated in which, with

$$J \equiv (j, j^*)(i/\hbar)$$

and

$$\Psi \equiv (\psi, \psi^+),$$

we have

$$\langle\Psi(1)\rangle = \frac{\delta W}{\delta J(1)} \qquad \text{(see fig. 1a)},$$

$$G(12) = \frac{\delta^2 W}{\delta J(1)\delta J(2)} = \langle(\Psi(1)\Psi(2))_+\rangle - \langle\Psi(1)\rangle\langle\Psi(2)\rangle \qquad \text{(see fig. 1b)}$$

as well as higher order cumulants or connected propagators defined by

$$G(1\ldots n) = \frac{\delta^n W}{\delta J(1)\cdots\delta J(n)}.$$

It is natural also to introduce vertex functions described in terms of the Legendre transform, X. Specifically, we have

$$\frac{\delta}{\delta\langle\Psi(1)\rangle}X \equiv \frac{\delta}{\delta\langle\Psi(1)\rangle}\left[-W + \int J(1)\Psi(1)\right] = J(1),$$

$$\Gamma(12) = G^{-1}(12) \equiv G_0^{-1}(12) - \Sigma(12) = \frac{\delta^2 X}{\delta\langle\Psi(1)\rangle\delta\langle\Psi(2)\rangle},$$

$$\Gamma(1\ldots n) = \frac{\delta^n X}{\delta\langle\Psi(1)\rangle\delta\langle\Psi(2)\rangle\cdots\delta\langle\Psi(n)\rangle}.$$

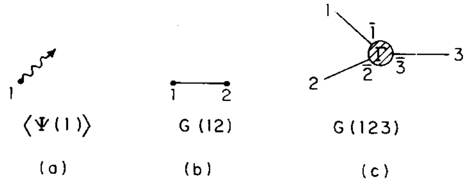


Fig. 1. Diagrammatic elements representing the mean value of the field and its cumulants. The circle represents the three point vertex.

With a summation convention for repeated indices, we may write

$$G(123) = G(1\bar{1})G(2\bar{2})G(3\bar{3})\Gamma(\bar{1}\bar{2}\bar{3}) \quad \text{(see fig. 1c).}$$

Higher order "real physical" interactions are similarly related to higher order cumulants. The development of self-consistent perturbation theories in terms of these true propagators and, if desired, in terms of true interactions, follows simply and mechanically from the basic equations.

Specifically, we obtain a self-energy

$$\Sigma(11') = \lambda \frac{\delta}{\delta\langle\Psi(1')\rangle}\left[G(12)\frac{\delta}{\delta\langle\Psi(2)\rangle} + \langle\Psi(1)\rangle\right] \times\left[G(13)\frac{\delta}{\delta\langle\Psi(3)\rangle} + \langle\Psi(1)\rangle\right]\langle\Psi(1)\rangle$$

which gives rise to the terms which are depicted in fig. 2.

From these equations the functional equations by which perturbation theory is generated, follow immediately. An uncondensed Bose system, in which $\langle\Psi(1)\rangle = 0$, appears as a special case in which the terms described by figs. 2a, 2c, and 2e vanish. By contrast, in a highly condensed system, for example, a weakly interacting Bose gas in its ground state, the term described by fig. 2a dominates the one in fig. 2b and the term in fig. 2c dominates those

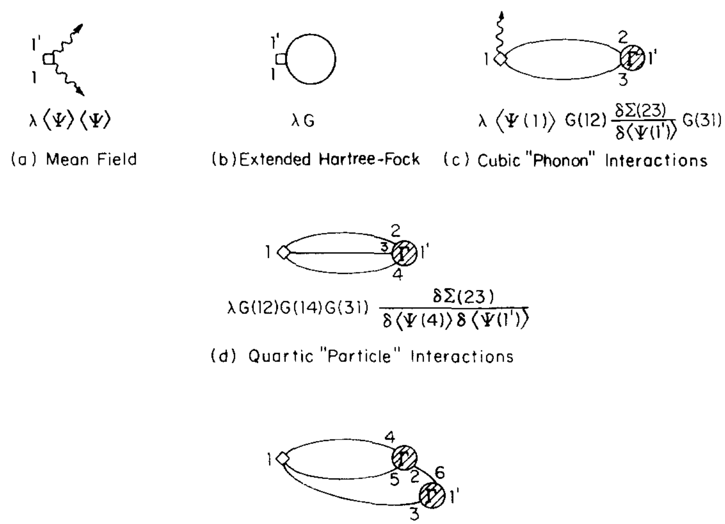


Fig. 2. Exact self-energy for a Bose propagator. The circles represent vertices and the square, the four point interaction.

in figs. 2d and 2e. With this approach, the condensed system in which $G(11)/|\langle\Psi(1)\rangle|^2 \ll 1$, occurs no less naturally than the "normal" uncondensed system.

In short, the perturbation framework developed by Julian is superior to the conventional scheme in that:

1) It allows for and "insists upon" the possibility for anomalous propagators. This possibility arises naturally because the theory is phrased entirely in terms of "true", rather than "bare", propagators.

2) It makes no "adiabatic" perturbative assumption, and thus allows naturally for self-consistent solutions.

3) At no stage does it entail unphysical "unlinked diagrams." Their absence does not rest on a "Wick theorem" (which does not hold for operators that do not satisfy canonical commutation relations).

In order to make the framework less schematic, it is necessary to impose boundary conditions. In particular:

1) To discuss the ground state of a relativistic field theory, the differential equations must be supplemented by positive frequency boundary conditions. These conditions, and how to incorporate them with a Euclidian formulation were explained by Julian in another cryptic article[4]).

2) To discuss systems in thermal equilibrium, the equations must be supplemented by[2]) a periodic boundary condition in imaginary time[3]). This condition, which is tantamount to the fluctuation-dissipation theorem, seems to have acquired a name, the KSM condition, in a literature that is by now inscrutable to Kubo, Martin, and Schringer.

3) To discuss non-equilibrium quantum systems, it is necessary to specify initial – not boundary conditions. The treatment of such quantum systems was first discussed by Julian[5]) in 1961. Without equilibrium, there is no general connection between fluctuation and dissipation, and as a result more independent functions must be determined. The subsequent developments of this approach[6]) by many authors are not always valid. Specifically, the usual treatments hold only for special initial conditions, or after the system has evolved for a time long enough to eliminate most of the dependence on initial conditions. (Naturally, the existence of such a time and its value can vary from one property to another and from one system to another. In addition, some aspects of these conditions, e.g., the total mass and energy, remain forever.) With these treatments steady states far from equilibrium, e.g., a continuously pumped laser, can be analyzed.

4) To discuss non-equilibrium classical systems, it is useful to introduce for each classical field ψ a second field, $\delta/\delta\psi$. In the terms of two fields the analysis[7]) of classical spins and of Navier–Stokes' fluids becomes simpler, clearer, and more systematic[8]).

5) Although a full treatment of non-equilibrium systems with arbitrary initial conditions by functional techniques has never been spelled out in the literature, it poses no problem. At short times one obtains complicated equations that describe the effects of both interactions and initial correlations in terms of "linked diagrams"[9]).

6) Finally, Julian and some of the rest of you may also be amused to hear of some non-esoteric complications the non-equilibrium formalism masks. We have recently discovered that the theory does not forget, although users might like to, the fact that non-equilibrium steady states cannot be maintained without carrying off the heat produced by the work performed by external forces. Spatial boundary conditions, and the heat exchange at spatial boundaries play a far greater role outside of equilibrium (and in nonlinear response) than they do in equilibrium (or linear response)[10]). One illustration of this complication is the divergence of the term of order E^2 of the current fluctuations of an electron gas with impurities when there is no heat sink.

3. A case study – superfluid helium

Below 2.2 K, liquid helium behaves very strangely: It is hard to contain, flowing through minute capillaries without friction; heat propagates through it like sound, and a temperature gradient produces a flow. To describe it phenomenologically, an extra variable, an irrotational "superfluid velocity," that does not appear in the Navier–Stokes' equations is needed. An additional equilibrium parameter, the "superfluid density" is also required. A two fluid model[11]) which allows for persistent flow was developed by Landau, who placed great stress on the elementary excitation spectrum of helium, with phonons and rotons (see fig. 3).

The Landau picture did not emphasize the Bose nature of ^4He and its relation to superfluidity. Indeed it seems that the Russian school anticipated that ^3He might also be superfluid. The importance of Bose condensation and its implications for macroscopic quantization of vorticity were first realized

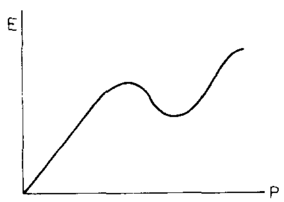


Fig. 3. Energy-momentum relation for excitations of superfluid helium as determined by neutron-scattering experiments.

by Onsager and emphasized by Feynman[12]) in the early 50's. The connection of this picture with states in which $\langle\Psi\rangle \neq 0$ seems first to have been noted in the work of Bogolyubov[13]). By the mid-1960's it was generally recognized that below the transition temperature, the phase of the condensate $\langle\Psi(r)\rangle \equiv \sqrt{n_0(r)}\exp(i\phi(r))$ should be associated with the irrotational superfluid velocity potential, that is,

$$v_s(r) = (\hbar/m)\nabla\phi(r) \tag{2.1}$$

and the fraction of particles in the macroscopically condensed mode with $n_0(r)/n$ [where n is the density of particles]. That $n_0(r)$ differs from the *superfluid density*, $n_s(r)$, measured in heat propagation and rotation experiments described by energy and current correlation functions, had been made clear by derivations of the two fluid model and other properties of helium[14]). These derivations identify many of the measurable parameters with the vertices described in the previous section of this paper. With condensation and quantization, superfluidity may be easily understood.

Far below T_c, it was possible to derive the properties described in fig. 4, and to understand why, although the Landau criterion, a necessary condition for superfluidity, could be illustrated in fig. 3, this criterion was less essential than condensation. The listener who finds this statement mysterious should reflect on the modification of the excitation spectra, low temperature specific heat, condensate fraction, and superfluid density of superfluid ^{4}He when it contains a small concentration of ^{3}He impurities.

The crosses on the n_s curves in fig. 4 are schematic. They are intended to show that n_s is easily measured but that n_0 is not. Indeed, although estimates for n_0/n at $T = 0$ of about $\sim .08$ have existed for some time, the first fairly reliable measurement of n_0 was made only last year. This measurement by Woods and Sears[15]) is represented by the single cross on the n_0/n curve.

To all orders in perturbation theory, for low frequencies, ω, wave numbers,

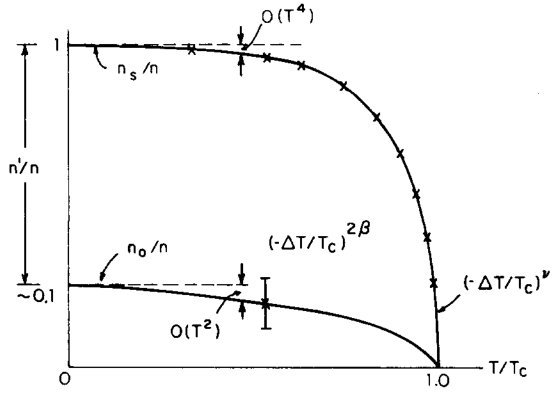


Fig. 4. Condensate fraction and superfluid fraction of helium as a function of temperature.

q, and temperatures, T, it is possible to prove that the Fourier transform of the field correlation function obeys the relation

$$\text{F.T.}\quad i\eta(t)\langle[\psi(rt),\psi^{+}(00)]\rangle \approx \frac{n_0 m^2 c^2}{n_s m h^2}\frac{1}{c^2q^2-\omega^2}, \tag{2.2}$$

where $c^2 = (\mathrm{d}p/\mathrm{d}n)$ and $n_s \to n$. Thus, at long wavelengths and low frequencies, the Bose gas has phonons. It is also possible to show that, for small momenta, p, the momentum distribution, $n(p)$ is given by

$$n(p) \cong n_0(T)\left[\delta(p) + \frac{k_B T m}{h^3 p^2 n_s(T)} + \frac{mc}{2h^3 np}\right], \tag{2.3}$$

$$n = \int \mathrm{d}^3 p\, n(p).$$

These important results can be qualitatively explained in the following manner. The classical equipartition theorem holds for long wavelength excitations and implies that

$$\tfrac{1}{2} m n_s \langle v_s^2\rangle \sim \tfrac{1}{2} k_B T. \tag{2.4}$$

The connection between v_s and the phase ϕ implies that

$$\frac{h^2}{m^2} m n_s q^2 \langle \phi_q^2\rangle = k_B T, \tag{2.5}$$

and the relation $\delta\psi \sim \sqrt{n_0}\, i\delta\phi$ yields

$$\langle \psi^{+}\psi\rangle_p \sim \frac{n_0}{n_s}\frac{k_B T m}{p^2(2\pi\hbar)^3}. \tag{2.6}$$

The connection between the mean classical energy $k_B T$ and the mean quantum energy $cp[(e^{cp/k_B T}-1)^{-1}+\frac{1}{2}]$ is the source of the last term in (2.3).

The phase fluctuations associated with the Goldstone mode, that must be present because when $\langle\psi\rangle = \sqrt{n_0}e^{i\phi} \neq 0$ states with differing phase ϕ are degenerate, have far-reaching consequences for condensed Bose systems.

For example, in terms of the "inverse stiffness"

$$K_s^{-1} \equiv k_B T m / h^2 n_s \tag{2.7}$$

which has the dimension of (length)$^{d-2}$ where d is the spatial dimension of the system, we may infer that[16])

$$\begin{aligned}\langle\psi^{+}(0)\psi(r)\rangle &\sim \sqrt{n_0(0)}\, e^{-i\phi(0)}\sqrt{n_0(r)}\, e^{i\phi(r)} \\ &\sim n_0\langle e^{-i\phi(0)}\, e^{i\phi(r)}\rangle \sim n_0\, e^{-\langle(\phi(0)-\phi(r))^2\rangle/2} \\ &\sim \exp\left[-\int \frac{d^d q}{(2\pi)^d}\frac{1}{K_s q^2}e^{iq\cdot r}\right]\end{aligned} \tag{2.8}$$

which yields for various spatial dimensions, d:

d	$\langle e^{i\phi(0)} e^{-i\phi(r)} \rangle$	Implication
1	$\exp[-K_s^{-1} r]$	No long range order
2	$(r)^{-(2\pi K_s)^{-1}}$	No long range order but infinite correlation range for low T ($K_s \to \infty$)
3	$1-\exp[-(4\pi K_s r)^{-1}]$	Long range order
4	$1-\exp[-(2\pi^2 K_s r^2)^{-1}]$	Long range order

(2.9)

These qualitative results are borne out by more careful investigations. The conclusions for ($d = 2$) are probably the most interesting since they imply that two dimensional Bose systems (i.e., helium films) undergo a phase transition to a state with an infinite correlation length but no long range order. We shall return to this conclusion after some brief comments on the behavior of bulk helium near T_c.

4. Systems with no simple counterpart

It is standard lore among physicists that the only exactly soluble problem is the harmonic oscillator (and by extension, the only approximately soluble problems are those that can be studied by perturbing the oscillator). Julian has done more than his share to contribute to this rule, reducing the study of angular momenta to the oscillator[16]) and treating the hydrogen atom and its Stark effect in terms of angular momenta[17]). Indeed the hand-waving generalist might claim that all the self-consistent theories discussed to this point in this lecture can be put into one-to-one correspondence with theories of weakly coupled oscillators (some of which are spontaneously displaced) and that the great advances in condensed matter physics in the past decade have come from understanding problems for which no such one-to-one mapping occurs.

In a very deep sense, the most amazing and striking features of continuous phase transitions *are* connected with the absence of such a one-to-one mapping onto a set of weakly interacting excitations. The behavior of classes of systems with different Hamiltonians becomes identical at the transition point *because*, in the asymptotic limit that characterizes the transition, a non-trivial non-oscillator like "fixed-point" Hamiltonian describes the dominant features. There are, of course, many allusions to fixed-points and to the disappearance of certain parameters in asymptotic limits throughout the field theory literature of the 1950's and 60's. Much occurs under the label, "renormalization group." However, our understanding of the behavior of physical systems at their transition temperatures, where even self-consistent perturbation theory totally fails, really commenced with the work of Leo Kadanoff[18]) in 1966, and the problem was essentially unravelled by Kenneth Wilson[19]) in 1971 and 1972.

I shall not attempt to summarize here the ideas behind "renormalization group" techniques that Wilson used to explain critical phenomena. They are discussed in ref. 19c. Let me mention, however, one relevant observation: Both Leo Kadanoff and Ken Wilson, who made those advances, were deeply imbued with Schwinger-style physics. Each did their undergraduate work and each spent several additional years at Harvard – Kadanoff as a graduate student, and Wilson as a member of the Society of Fellows.

Suffice it to say that through their work and the work of many others we know that Bose systems which undergo continuous phase transitions to condensed states have singular properties at T_c which depend *only on the dimension*, d; they are completely independent of the interaction. Moreover, essentially all the leading asymptotic measurable properties of these systems can be characterized in terms of two accurately calculable numbers, the critical exponents, β and ν, defined in fig. 4. The best value[20]) for $\nu \simeq 0.670$ has been verified rather precisely. The prediction, $\beta \simeq 0.346$, is more difficult to study experimentally for the reason noted earlier. β is related to the parameter, η, defined in terms of the asymptotic behavior for small k of the Fourier transform of the order parameter correlation function $G(k) \sim k^{-2+\eta}$ by

$$\eta = 2 - d + 2\beta/\nu = 0.0335. \tag{3.1}$$

In field theories where the broken symmetry is described not by a single angle ϕ, but by a symmetry group $0(n)$, with $n \neq 1$, the critical exponents have different values. In particular, theories in which $n = 3$, 4, etc. with non-Abelian symmetry groups, have no phase transitions in two dimensions[21]).

Because there is a close relation between the properties of field theories for Bose particle fields with $0(n)$ symmetry in two dimensions and lattice gauge field theories with $0(n)$ symmetry in four dimensions, the asymptotic behavior of a two-dimensional Bose gas and of quantum electrodynamics on a lattice are closely connected. Specifically the fact that correlations in a Bose gas fall off algebraically at low temperatures and exponentially at high temperatures is related to the fact that on a lattice, electrodynamics results in massless photons when the coupling is weak but to "confinement" when the coupling is strong. Likewise, for $n > 2$ there are no "low-temperature" long range correlations in two spatial dimensions, and when $n > 2$ in a non-Abelian gauge field theory there may always be "confinement." This connection, which has been hypothesized and examined by Wilson, Migdal[21]), and Polyakov[22]) is

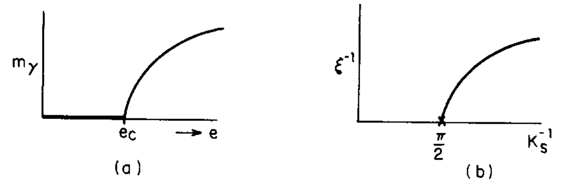


Fig. 5. Parallel plots of the mass of "photon" excitations as a function of the coupling constant and of the inverse coherence length of two-dimensional helium as a function of the temperature.

discussed at some length in a recent review by Kadanoff[25]). The picture in fig. 5, describing the connection, is contained in Wilson's 1974 article. It may remind Julian of a paper[26]) he wrote fifteen years ago.

While these arguments make the two-dimensional Bose gas below T_c a system worthy of study not only by those concerned with helium films and similar materials, a full discussion of the results of Kosterlitz and Thouless[27]) and of José, Kadanoff, Kirkpatrick, and Nelson[28]) cannot be carried out here. I would, however, like to describe in some detail one striking and directly verifiable prediction Nelson and Kosterlitz discussed in a recent letter[29]): n_s does not vanish for $T < T_c$ in two dimensions even though n_0 does. In fact, n_s/T or K_s is not only finite and nonvanishing below T_c but it remains nonvanishing, decreasing to the value $2/\pi$ at T_c and then dropping abruptly to zero. This value is *universal* even though T_c and n_s vary with the thickness of the film. Steps to test this prediction are currently underway at Cornell[30]).

As a result of this talk at UCLA, I. Rudnick was made aware of the predictions in this paragraph. Immediately afterward he undertook a re-analysis of his earlier data which had never been compiled in this fashion, and discovered that the agreement was excellent[31]). *Even if this review has no other consequences, it can lay claim to one very practical contribution to scientific communication.*

Nelson and Kosterlitz begin with the Hamiltonian

$$\frac{H}{kT} = \int \mathrm{d}r \left[\tfrac{1}{2}K\left(\frac{m}{\hbar}\, v_s\right)^2 - m^2 \ln y_0 |\nabla \times v_s|^2/(2\pi\hbar)^2\right] \tag{3.2}$$

in which K is a "bare" stiffness constant. The second term in this Hamiltonian describes the vorticity and is multiplied by a "bare" vorticity chemical potential, $\ln y_0$. The superfluid velocity, in turn, is described by

$$v_s = \frac{\hbar}{m}\nabla\phi + \frac{2\pi\hbar}{m}(\hat{z}\times\nabla)\int \mathrm{d}^2 r'\,\omega(r')\left\langle r\left| -\frac{1}{\nabla^2}\right| r'\right\rangle, \tag{3.3}$$

where ω is an integral-valued vorticity field. They rewrite the Hamiltonian in the form

$$\frac{H}{kT} \sim \tfrac{1}{2} K \int \mathrm{d}^2 r (\nabla\phi)^2 - \pi K \int \frac{\mathrm{d}^2 r}{a^2} \frac{\mathrm{d}^2 r'}{a^2} \omega(r)\omega(r') \ln \frac{|r - r'|}{a} - \ln y \int \frac{\mathrm{d}^2 r}{a^2} \omega^2(r). \tag{3.4}$$

where a, a vortex core radius has been explicitly inserted.

Clearly, when $\int \langle \omega(r) \rangle \, \mathrm{d}^2 r = 0$, the Hamiltonian is the same as that for a two-dimensional neutral Coulomb gas (which has finite energy). The renormalized stiffness constant or n_s is rigorously given by

$$K_s^{-1} = \frac{m^2}{\hbar^2} \int \mathrm{d}^2 r \langle v_s(r) v_s(0) \rangle. \tag{3.5}$$

The vortex correlation function in eq. (2.13) $\langle \omega(r)\omega(0) \rangle$ is given approximately by

$$\langle \dot{\omega}(r)\omega(0) \rangle \sim -2y^2 \exp(-2\pi K \log r). \tag{3.6}$$

The correlation is attractive; it is proportional to the square of the vortex density or concentration; there are two orientations for each pair; and it varies as the exponential of the intervortex force. From eqs. (3.3), (3.5), and (3.6) they conclude that

$$K_s^{-1} = K_{(\nabla\phi)^2}^{-1} + 4\pi^3 y^2 \int_a^\infty \frac{\mathrm{d}r}{a} \left(\frac{r}{a}\right)^{3-2K\pi} + \mathcal{O}(y^4). \tag{3.7}$$

The implications of this equation are conveniently studied by renormalization group techniques. Specifically in this way the equation can be shown to describe the fact that as the effective length scale, l, is increased

$$\begin{aligned} \frac{\mathrm{d}y(l)}{\mathrm{d}l} &= y(l)[2 - \pi K(l)], \\ \frac{\mathrm{d}K^{-1}(l)}{\mathrm{d}l} &= 4\pi^3 y^2(l). \end{aligned} \tag{3.8}$$

The system is stable, and the vortex concentration decreases with increasing l and the integral in eq. (2.16) is convergent when $4 - 2K\pi < 0$; the system is unstable to vortex production when $K^{-1} > \pi/2$!!

The flows associated with eq. (3.8) are depicted in fig. 6. From eqs. (2.9) and (3.6) it follows that the field correlations reflecting the spin waves,

$$\langle \psi^+ \psi \rangle \sim r^{-(2\pi K_s)^{-1}} \tag{3.9}$$

die off slowly at low T, while the vorticity correlations, reflecting the bound vortex pairs,

$$\langle \omega\omega \rangle \sim r^{-2K_s\pi} \tag{3.10}$$

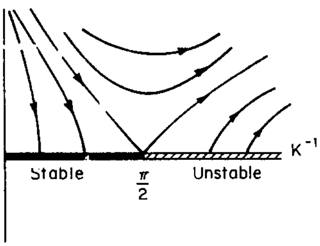


Fig. 6. Flows of the vortex chemical potential and the temperature as the length scale increases.

die off rapidly. These correlation ranges shrink and grow, respectively, and, at the critical temperature, with $K = 2/\pi$,

$$\langle \psi^+ \psi \rangle_{T_c} \sim r^{-1/4}, \qquad \langle \omega\omega \rangle_{T_c} \sim r^{-4}. \tag{3.11}$$

A table giving the values of the critical exponents, η and ν, as a function of dimension for systems with discrete ($n = 1$) symmetry (e.g., the Ising model or a gas–liquid transition); Abelian ($n = 2$) continuous symmetry (e.g., our Bose gas system or a planar ferromagnet); and one non-Abelian ($n = 3$) symmetry follows:

$\xi \equiv (\text{mass})^{-1} \sim (\Delta T)^{-\nu}$ $\langle \psi^+(r)\psi(0) \rangle \sim r^{-d+2-\eta}$

$n = 1$

d	ν	η
4	$\frac{1}{2}$	0
3	$\approx .630$	$\approx .031$
2	1	$\frac{1}{4}$
1	no transition	

$n = 2$

d	ν	η
4	$\frac{1}{2}$	0
3	$\approx .670$	$\approx .033$
2	∞*	$\frac{1}{4}$

*$\xi \sim \exp[c(\Delta T)^{-1/2}]$

$n = 3$

d	ν	η
4	$\frac{1}{2}$	0
3	$\approx .705$	$\approx .034$
2	no transition	

In support of my claim that condensed matter offers unparalleled opportunities for studying field theories, I cannot resist pointing out that condensed matter theorists have even found a way to enter and experiment in the fourth dimension. Experiments on $LiTbF_4$ test[32]) the results for $d = 4$ where the asymptotic corrections to free fields are logarithmic.

The results discussed to this point in this section are intrinsically non-perturbative. Nevertheless, for technical reasons connected with approximations in a variable dimension many studies of the asymptotic properties of the states onto which systems map at the critical point have made use of many perturbative field theoretic techniques. (Others, for example, the "real-space" methods, have not.) As a result, critical phenomena have provided many-particle physicists with a set of new but not completely foreign challenges. In the remainder of this section, I would like to turn to a problem that I fear offers a more formidable challenge. That is the problem of chaos or turbulence. On this matter I like to recall the statement[33]) attributed to Sir Horace Lamb in 1932:

> "I am an old man, and when I die and go to heaven, there are two matters on which I hope for enlightenment. One is quantum electrodynamics and the other is the turbulent motions of fluids. About the former I am really rather optimistic."

Viscous fluids at rest, or subjected to weak mechanical or thermal stresses relax to a well defined state when they are disturbed from equilibrium. In this respect, they and simpler mechanical dissipative systems behave in time, in a manner similar to the way in which coupling constants flow when the scale is changed. They approach fixed points. When the external stress is sufficiently large, however, the opposite phenomenon occurs. The "fixed point" or time-independent solution becomes unstable and the system exhibits erratic time-varying behavior though there are no time-varying external parameters. Exactly how the system behaves depends on the precise initial conditions and only statistical aspects of the behavior are predictable. Systems with very similar initial conditions behave entirely differently at much later times. It is not unreasonable to say that such systems are characterized by a "one-to-many" mapping – a mapping which is considerably more perverse than even the "many-to-one" mappings that characterize fixed points[34]).

Not surprisingly most of the research by physicists on the properties of turbulence has been directed at the statistical properties of strongly and randomly forced fluid systems. In such systems there is a one-to-one mapping between a noisy stirring force and a noisy fluid spectrum. However, near the onset of turbulence, the internal noise that develops does *not* depend upon the external noise and even in fully developed turbulence (which is difficult to produce in the laboratory) the equivalence of the short distance internal fluctuation spectra produced by strong 'coherent" (noise-free) stresses and random (noisy) stresses is based largely on faith.

In order to make this point let me turn from the sublime to the concrete and call your attention to the simple mechanical system described in fig. 7 and actually constructed some years ago by Louis Howard and Willem Malkus[35]) at M.I.T.

The equations this simple mechanical system obeys were first analyzed in a very beautiful paper by Ed Lorenz[36]) that did not receive the attention it merits for many years. He introduced these equations after studying certain

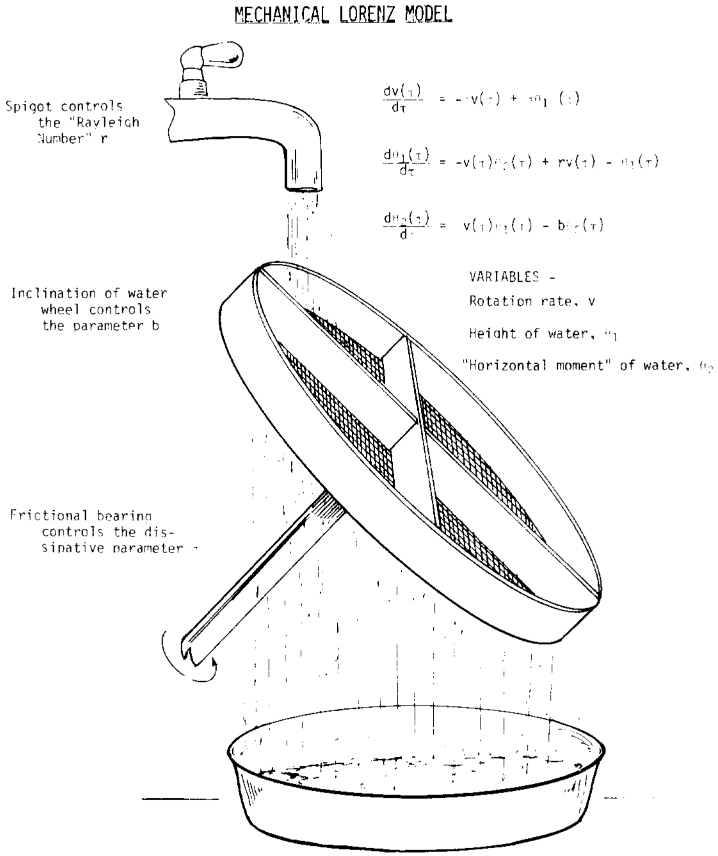


Fig. 7. A simple mechanical model that exhibits chaotic behavior.

numerical solutions to a truncated set of equations for a convecting fluid. Similar equations are now believed to play a role in modelling the earth's magnetic field and its reversals. Very briefly, when the water flows at a slow steady rate into the partitioned water wheel with a screened base, it leaks through the screen and, because the bearing is frictional, the wheel does not turn. When the rate, *r*, is increased to a value larger than unity, the weight of the water at the top of the tilted wheel is sufficient to overcome the friction and the wheel turns at a uniform rate either clockwise or counterclockwise ("convection"). Finally, when the water enters steadily but at a sufficiently fast rate, the wheel may turn too rapidly; the filled (heavy) partition of the wheel may then "overshoot" the bottom and rise partially, the unfilled (light) partition directly opposite it having insufficient time under the faucet to acquire sufficient counterweight. The wheel then turns in the opposite direction, moving rapidly to a different position, where, once again a reversal takes place. Lorenz showed that the dynamics of this system is, in a deep sense, chaotic. Although unpredictable in detail, the statistics of the motion are well defined and measurable, and much can be said about them. In recent years mathematicians and physicists have begun to study the properties of such equations. Using light scattering techniques that were developed and employed to study dynamic critical phenomena, physicists have begun to measure carefully the noise spectra of real fluid systems that undergo transitions to periodic and chaotic flows[37]). Exactly how much that is generally applicable can be learned from these studies is hard to say. In any event, I think you will find the wealth of information contained in the Lorentz equations thought-provoking and urge you to check that one stationary solution

$$v = 0, \quad \theta_1 = 0, \quad \theta_2 = 0$$

is stable only for $0 < r < 1$; to verify that a second

$$v = \theta_1 = \pm\sqrt{b(r-1)},$$
$$\theta_2 = r - 1$$

is stable only for

$$1 < r < r_T = \frac{\sigma(\sigma + b + 3)}{\sigma - b - 1},$$

the marginally stable modes at r_T having frequencies,

$$\omega_T = \pm\left[\frac{2b\sigma(\sigma+1)}{\sigma - b - 1}\right]^{1/2};$$

and then to look at Lorenz's paper to find out about the chaotic behavior that occurs for a range of values above r_T although not at all very large values of r.

In this talk I have probably undertaken too large a task. I have tried to review for Julian and for an audience comprised dominantly of field theorists how some of the ideas and methods to which Julian may lay claim have been extended in the field of condensed matter physics and how that field has evolved. I have tried to stress recent developments in directions of particular interest to this audience. The liquid helium properties I discussed exhibit one of several areas in which a confluence of concepts from particle and condensed matter physics has recently occurred. Others include topological singularities (from dislocations and textures to monopoles and instantons) and exactly soluble models (from the Luttinger and Luther models to the Sine–Gordon equation and Thirring model). Finally, I have briefly discussed an unresolved fascinating problem and tried to explain why it is difficult. Experience suggests that progress on this problem may also be widely applicable in field theory but it is so different from the other problems we have resolved that the potential connections are difficult to visualize[38]).

Above all I have tried to document the profound influence that Julian Schwinger has had, through his teaching as well as his papers on one of the major fields of physics with which his name is not ordinarily associated. I hope I have done justice to the case.

Acknowledgement

I would like to thank David Nelson for many helpful comments.

References

1) J. Schwinger, Proc. Nat. Acad. Sci. **37** (1951) 452.
2) J. Schwinger, Phys. Rev. **82** (1951) 914; **91** (1953) 713; **91** (1953) 728; **92** (1953) 1283; **93** (1954) 615; **94** (1954) 1362.
3) P.C. Martin and J. Schwinger, Phys. Rev. **115** (1959) 1342.
4) J. Schwinger, Proc. Nat. Acad. Sci. **44** (1958) 956;
See also J. Schwinger, Phys. Rev. **115** (1959) 721 and Phys. Rev. **117** (1960) 1407.
5) J. Schwinger, J. Math. Phys. **2** (1961) 407.
6) L.V. Keldysh, Sov. Phys. JETP **20** (1965) 1018.
See also D.C. Langreth, 1975 NATO Advanced Study Institute on Linear and Non-Linear Electron Transport in Solids (Plenum, Antwerp, 1976), p. 3.
7) Classical perturbation theory in fluids was discussed by H. W. Wyld, Jr., Ann. Phys. **14** (1961) 143 and by R.H. Kraichnan, J. Fluid. Mech. **5** (1959) 497; J. Math. Phys. **2** (1961) 124; J. Math. Phys. **3** (1962) 205; Phys. Fluids **7** (1964) 1723.
8) P.C. Martin, E.D. Siggia and H.A. Rose, Phys. Rev. **A8** (1973) 423;
C. DeDominicis, J. de Phys. (Paris) **C1** (1976) 247;
H.K. Janssen, Z. Phys. **B23** (1976) 377;
R. Bausch, H.K. Janssen and H. Wagner, Z. Phys. **B24** (1976) 113.
9) See, for example, A.G. Hall, J. Phys. **A8** (1975) 214;
S. Fujita, Phys. Rev. **A4 (1971) 1114.**
10) A.-M. Tremblay, B. Patton, P.C. Martin and P. Maldague, "Microscopic Calculation of the Non-Linear Current Fluctuations of a Metallic Resistor: The Problem of Heating in Perturbation Theory" (to be published, Phys. Rev. A, April 1979).
11) L.D. Landau, J. Phys. (USSR) **5** (1941) 71; **11** (1947) 91.
12) R.P. Feynman, Progress in Low Temperature Physics, Vol. I, Chap. 11 (North-Holland, Amsterdam, 1957).
13) N. Bogolyubov, J. Phys. (USSR) **11** (1947) 23.
14) P.C. Hohenberg and P.C. Martin, Ann. of Phys. (NY) **34** (1965) 291 where many previous references are cited;
W. Kane and L.P. Kadanoff, Phys. Rev. **155** (1967) 80;
P.C. Martin, in Statistical Mechanics at the Turn of the Decade, E.G.D. Cohen, ed. (Marcel Dekker, New York, 1971).
15) A.D.B. Woods and V.F. Sears, PRL **39** (1977) 415.
16) J. Schwinger, on Angular Momentum (1952). Published in Quantum Theory of Angular Momentum (Academic Press, New York, 1965).

17) J. Schwinger, unpublished lecture notes on Quantum Mechanics, Harvard University (1952). See Wu-yang Tsai, Phys. Rev. **A9** (1974) 1081.
18) L.P. Kadanoff, Physics **2** (1966) 263; L.P. Kadanoff, et al., Rev. Mod. Phys. **39** (1967) 395.
19) K.G. Wilson and M.E. Fisher, PRL **28** (1972) 240;
K.G. Wilson, PRL **28** (1972) 548;
K.G. Wilson and J. Kogut, Phys. Rep. **12C** (1974) 77.
20) J. LeGuillou and J. Zinn-Justin, PRL **39** (1977) 95.
21) V.L. Berezinskii, Sov. Phys. JETP **32** (1971) 493; **34** (1972) 610.
See also E. Brézin and J. Zinn-Justin PRL **36** (1976) 691.
22) K.G. Wilson, Phys. Rep. **23C** (1976) 331; **D10** (1974) 2445.
23) A.A. Migdal, Sov. Phys. JETP **42** (1975) 413, 743.
24) A.M. Polyakov, Phys. Lett. **B59** (1975) 79.
25) L.P. Kadanoff, Rev. Mod. Phys. **49** (1977) 267.
26) J. Schwinger, Phys. Rev. **128** (1962) 2425.
27) J.M. Kosterlitz and D.J. Thouless, J. Phys. **C5** (1972) L124; **C6** (1973) 1181.
28) J.V. José, L.P. Kadanoff, S. Kirkpatrick and D. Nelson, Phys. Rev. **B16** (1977) 1217.
29) D.R. Nelson and J.M. Kosterlitz, PRL **39** (1977) 1201.
30) D.J. Bishop and J.D. Reppy, PRL **40** (1978) 1727.
31) I. Rudnick, PRL **40** (1978) 1454.
32) G. Ahlers, A. Kornblit and H. Guggenheim, PRL **34** (1975) 513.
For the theory, see A.I. Larkin and D.E. Khmel'nitskii, Soviet Phys. JETP **29** (1969) 1123 and E. Brézin and J. Zinn-Justin, Phys. Rev. **B13** (1976) 251.
33) S. Goldstein, Fluid Mechanics in the First Half of this Century, Ann. Rev. Fl. Mech. I (1969) 23.
34) To be slightly more precise, we need a tractable model or models (if they exist), onto which real chaotic systems can be perturbatively mapped much as real condensed systems are mapped against the "many" degenerate mean field states that develop from a single uncondensed state as the temperature decreases.
35) L.N. Howard and W.V.R. Malkus (private demonstration). The model is briefly mentioned in W.V.R. Malkus, Mémoires Société Royale des Sciences de Liège, 6th Series, Vol. 4, (1972) 125.
36) E.N. Lorenz, J. Atmos. Sci. **20** (1963) 130.
37) See, for example, the recent popular review by H.L. Swinney and J.P. Gollub, Physics Today **31** (1978) 41.
38) One possible connection might be between the coupling constant flows of the renormalization group analysis and the phase space flows of the dynamical variable of the chaotic dissipative mechanical system. In the limit of fully developed turbulence there is another obvious possibility discussed, for example, by M. Nelkin, in Proceedings of STATPHYS **13**, IUPAP Conference on Statistical Physics, Haifa, Israel, August 1977. See also M. Nelkin and T.L. Bell, Phys. Rev. **A17** (1978) 363 and P.C. Martin and C. DeDominicis, The Long Distance Behavior of Randomly Stirred Fluids, Supp. of Prog. Theor. Phys. **64** (1978).

Slow Electrons in a Polar Crystal

R. P. FEYNMAN
California Institute of Technology, Pasadena, California
(Received October 19, 1954)

A variational principle is developed for the lowest energy of a system described by a path integral. It is applied to the problem of the interaction of an electron with a polarizable lattice, as idealized by Fröhlich. The motion of the electron, after the phonons of the lattice field are eliminated, is described as a path integral. The variational method applied to this gives an energy for all values of the coupling constant. It is at least as accurate as previously known results. The effective mass of the electron is also calculated, but the accuracy here is difficult to judge.

AN electron in an ionic crystal polarizes the lattice in its neighborhood. This interaction changes the energy of the electron. Furthermore, when the electron moves the polarization state must move with it. An electron moving with its accompanying distortion of the lattice has sometimes been called a polaron. It has an effective mass higher than that of the electron. We wish to compute the energy and effective mass of such an electron. A summary giving the present state of this problem has been given by Fröhlich.[1] He makes simplifying assumptions, such that the crystal lattice acts much like a dielectric medium, and that all the important phonon waves have the same frequency. We will not discuss the validity of these assumptions here, but will consider the problem described by Fröhlich as simply a mathematical problem. Aside from its intrinsic interest, the problem is a much simplified analog of those which occur in the conventional meson theory when perturbation theory is inadequate. The method we shall use to solve the polaron problem is new, but the pseudoscalar symmetric meson field problems involve so many further complications that it cannot be directly applied there without further development.

We shall show how the variational technique which is so successful in ordinary quantum mechanics can be extended to integrals over trajectories.

STATEMENT OF THE PROBLEM

With Fröhlich's assumptions, the problem is reduced to that of finding the properties of the following Hamiltonian:

$$H=\tfrac{1}{2}\mathbf{P}^2+\sum_K a_K^+a_K+i(\sqrt{2}\pi\alpha/V)^{\frac{1}{2}}\sum_K \frac{1}{K}\times[a_K^+\exp(-i\mathbf{K}\cdot\mathbf{X})-a_K\exp(i\mathbf{K}\cdot\mathbf{X})]. \quad (1)$$

Here $\mathbf{X}$ is the vector position of the electron, $\mathbf{P}$ its conjugate momentum, a_K^+, a_K the creation and annihilation operators of a phonon (of momentum $\mathbf{K}$). The frequency of a phonon is taken to be independent of $\mathbf{K}$. Our units are such that $\hbar$, this frequency, and the electron mass are unity. The quantity α acts as a coupling constant, which may be large or small. In conventional units it is given by

$$\alpha=\frac{1}{2}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right)\frac{e^2}{\hbar\omega}\left(\frac{2m\omega}{\hbar}\right)^{\frac{1}{2}},$$

where ϵ, ϵ_∞ are the static and high frequency dielectric constant, respectively. In a typical case, such as NaCl, α may be about 5. The wave function of the system satisfies ($\hbar=1$)

$$i\partial\psi/\partial t=H\psi, \quad (2)$$

so that if φ_n and E_n are the eigenfunctions and eigenvalues of H,

$$H\varphi_n=E_n\varphi_n, \quad (3)$$

then any solution of (2) is of the form

$$\psi=\sum_n C_n\varphi_n e^{-iE_nt}.$$

Now we can cast (1) and (2) into the Lagrangian form of quantum mechanics and then eliminate the field oscillators (specializing to the case that all phonons are virtual). Doing this in exact analogy to quantum electrodynamics,[2] we find that we must study the sum over all trajectories $\mathbf{X}(t)$ of $\exp(iS')$, where

$$S'=\frac{1}{2}\int\left(\frac{d\mathbf{X}}{dt}\right)^2dt+2^{-\frac{1}{2}}\alpha i\int\int|\mathbf{X}_t-\mathbf{X}_s|^{-1}e^{-i|t-s|}dtds. \quad (4)$$

This sum will depend on the initial and final conditions and on the time interval T. Since it is a solution of the Schrödinger Eq. (2), considered as a function of T it will contain frequencies E_n, the lowest of which we seek. It is difficult to isolate the lowest frequency, however.

For that reason, consider the mathematical problem of solving

$$\partial\psi/\partial t=-H\psi, \quad (5)$$

without question as to the meaning of t. This has the same eigenvalues and eigenfunctions as (3), but a

[1] H. Fröhlich, Advances in Physics **3**, 325 (1954). References to other work is given here.

[2] R. P. Feynman, Phys. Rev. **80**, 440 (1950).

solution will have the form

$$\psi=\sum_n C_n e^{-E_n t}\varphi_n.$$

For large t any solution therefore asymptotically dies out exponentially, the last exponent surviving being that of the lowest E, say E_0.

An equation such as (5) can be converted to a path integral just as easily as (2) is, and the integral over the oscillator coordinates can again be done in an analogous way. The Lagrangian form corresponding to (5) turns out to be

$$K=\int \exp S\mathfrak{D}\mathbf{X}(t), \qquad (6)$$

with

$$S=-\frac{1}{2}\int\left(\frac{d\mathbf{X}}{dt}\right)^2 dt +2^{-\frac{3}{2}}\alpha\int\int|\mathbf{X}_t-\mathbf{X}_s|^{-1}e^{-|t-s|}dtds. \qquad (7)$$

This is just as one might expect from replacing t in (4) by $-it$. Now, since K is a solution of (5), its asymptotic form for a large t interval, 0 to T is

$$K\sim e^{-E_0 T} \qquad (8)$$

as $T\rightarrow\infty$. Therefore, we must estimate the path integral (6) for large T.

VARIATIONAL PRINCIPLE

The method we shall use is a type of variational method. Choose any S_1 which is simple and purports to be some sort of approximation to S. Then write

$$\int \exp S\mathfrak{D}\mathbf{X}(t)=\int \exp(S-S_1)\exp S_1\mathfrak{D}\mathbf{X}(t). \qquad (9)$$

Now this last expression can be looked upon as the average of $\exp(S-S_1)$, the average being taken with positive weight $\exp S_1$. But for any set of real quantities f the average of $\exp f$ exceeds the exponential of the average,

$$\langle\exp f\rangle\geq\exp\langle f\rangle. \qquad (10)$$

Hence if in (9) we replace $S-S_1$ by its average,

$$\langle S-S_1\rangle =\int(S-S_1)\exp S_1\mathfrak{D}\mathbf{X}(t)\Big/\int\exp S_1\mathfrak{D}\mathbf{X}(t), \qquad (11)$$

we will underestimate the value of (9). Therefore, if E is computed from

$$\int\exp(\langle S-S_1\rangle)\exp S_1\mathfrak{D}\mathbf{X}(t)\sim\exp -ET, \qquad (12)$$

then we know that E exceeds the true E_0,

$$E\geq E_0. \qquad (13)$$

If there are any free parameters in S_1 we can choose as the "best" values those which minimize E.

Since $\langle S-S_1\rangle$ defined in (11) is proportional to T, let us write

$$\langle S-S_1\rangle=sT. \qquad (14)$$

Furthermore, the factor $\exp\langle S-S_1\rangle$ in (12) is constant, of course, and may be taken outside the integral. Finally, suppose the lowest energy E_1 for the action S_1 is known,

$$\int\exp S_1\mathfrak{D}\mathbf{X}(t)\sim\exp(-E_1T), \qquad (15)$$

then we have

$$E=E_1-s \qquad (16)$$

from (12), with s given by (11) and (14). (In the case that S and S_1 are both simple actions [of the form of (18) below] this can readily be shown to be equivalent to the usual variational principle.)

POSSIBLE TRIAL ACTIONS

Some of the methods which have been applied to this problem, so far, correspond to various choices for S_1. The perturbation method corresponds to $S_1=-\frac{1}{2}\int(d\mathbf{X}/dt)^2dt$ and gives

$$E=-\alpha. \qquad (17)$$

We see immediately that the perturbation result is an upper limit to E_0, a result proven only with much greater effort by more usual methods, by Gurari[3] and Lee and Pines.[4] Another suggestion is

$$S_1=-\tfrac{1}{2}\int(d\mathbf{X}/dt)^2dt+\int V(\mathbf{X}_t)dt, \qquad (18)$$

where V is a potential to be chosen. If a Coulomb potential is chosen, $V(R)=Z/R$, and the parameter Z varied, one finds

$$E=-(25/256)\alpha^2=-0.098\alpha^2$$

asymptotically for the case that α is very large. For large α this corresponds to Landau's method[1] with a trial function of the form $e^{-\beta r}$. If a harmonic potential $V(R)=kR^2$ is used (corresponding to a Gaussian trial function in Landau's method) the value is somewhat improved:

$$E=-(1/3\pi)\alpha^2=-0.106\alpha^2. \qquad (19)$$

If α is not so large, the form (18) can still be used in (16). The evaluation of s requires knowledge of the eigenfunctions and eigenvalues for the potential V.

[3] M. Gurari, Phil. Mag. **44**, 329 (1953).
[4] T. Lee and D. Pines, Phys. Rev. **88**, 960 (1952). Lee, Low, and Pines, Phys. Rev. **90**, 297 (1953).

The result is somewhat difficult to evaluate for the Coulomb potential, but fairly simple for the harmonic case [see (34) below]. However, it is readily shown that for any α less than about 6 no choice of V can improve the result (17) for $V=0$. Fröhlich has asked for a method which works uniformly over the entire range of α. He points out that the artificial binding to a special origin, which (18) implies, is a disadvantage. It is this which presumably makes any potential V give a poorer result than $V=0$ for small α.

To remedy this, I thought a good idea would be to use for S_1 the action for a particle bound by a potential $V(\mathbf{X}-\mathbf{Y})$ to another particle of coordinate $\mathbf{Y}$. This latter could have finite mass, so no permanent origin would be assumed. Of course the action for such a system would contain both $\mathbf{X}(t)$ and $\mathbf{Y}(t)$. But the variables $\mathbf{Y}(t)$ could be integrated out, at least in principle, leaving an effective S_1 depending only on $\mathbf{X}$. At first I tried a Coulomb interaction for $V(\mathbf{X}-\mathbf{Y})$ but it was rather complicated. The technique may be useful in more difficult problems. But here we have already seen that an harmonic binding should be as good, if not better. Further, an extra particle bound harmonically has its variables $\mathbf{Y}(t)$ appearing quadratically in the action. It may therefore be easily eliminated explicitly. The result we know from studies of similar problems in electrodynamics. We are, in this way, led to consider the choice

$$S_1=-\tfrac{1}{2}\int\left(\frac{d\mathbf{X}}{dt}\right)^2 dt-\tfrac{1}{2}C\int\int[\mathbf{X}_t-\mathbf{X}_s]^2 \times\exp(-w|t-s|)dtds, \quad (20)$$

where C and w are parameters, to be chosen later to minimize E.

EVALUATION OF THE ENERGY

Since S_1 contains $\mathbf{X}$ only quadratically, all the necessary path integrals are easily done.[5] Because the method may not be familiar we outline it briefly here. Define the symbol $\langle\ \rangle$ as

$$\langle F\rangle=\int F\exp S_1\mathfrak{D}\mathbf{X}(t)\Big/\int\exp S_1\mathfrak{D}\mathbf{X}(t).$$

Then comparison of S_1 and S shows that

$$s=\frac{1}{T}\langle S-S_1\rangle=2^{-\frac{3}{2}}\alpha\int\langle|\mathbf{X}_t-\mathbf{X}_s|^{-1}\rangle e^{-|t-s|}ds +\tfrac{1}{2}C\int\langle(\mathbf{X}_t-\mathbf{X}_s)^2\rangle e^{-w|t-s|}ds=A+B. \quad (21)$$

[5] R. P. Feynman, Phys. Rev. 84, 108 (1951), Appendix C.

We concentrate first on the first term A of (21). In it we may express $|\mathbf{X}_t-\mathbf{X}_s|^{-1}$ by a Fourier transform,

$$|\mathbf{X}_t-\mathbf{X}_s|^{-1}=\int d^3\mathbf{K}\exp[i\mathbf{K}\cdot(\mathbf{X}_t-\mathbf{X}_s)](2\pi^2K^2)^{-1}. \quad (23)$$

For this reason we need to study

$$\langle\exp[i\mathbf{K}\cdot(\mathbf{X}_\tau-\mathbf{X}_\sigma)]\rangle =\int\exp S_1\exp[i\mathbf{K}\cdot(\mathbf{X}_\tau-\mathbf{X}_\sigma)]\mathfrak{D}\mathbf{X}(t)\Big/\int\exp S_1\mathfrak{D}\mathbf{X}(t). \quad (23)$$

The integral in the numerator is of the form

$$I=\int\exp\left[-\tfrac{1}{2}\int\left(\frac{d\mathbf{X}}{dt}\right)^2dt-\tfrac{1}{2}C\int\int(\mathbf{X}_t-\mathbf{X}_s)^2 \times e^{-w|t-s|}dtds+\int\mathbf{f}(t)\cdot\mathbf{X}(t)dt\right]\mathfrak{D}\mathbf{X}(t), \quad (24)$$

where specifically

$$\mathbf{f}(t)=i\mathbf{K}\delta(t-\tau)-i\mathbf{K}\delta(t-\sigma). \quad (25)$$

Now we shall find (24) insofar as it depends on $\mathbf{f}$ or $\mathbf{K}$ aside from a normalization factor which drops out in (23). Incidentally let us notice that the three rectangular components separate in (24) and we need only consider a scalar case. The method of integration is to substitute $X(t)=X'(t)+Y(t)$, where $X'(t)$ is that special function for which the exponent is maximum. The variable of integration is now $Y(t)$. Since the exponent is quadratic in $X(t)$ and X' renders it an extremum, it can contain $Y(t)$ only quadratically. Evidently Y then separates off as a factor not containing f, which may be integrated to give an unimportant constant (depends on T only). Therefore within such a constant

$$I=\exp\left[-\tfrac{1}{2}\int\dot{X}_t'^2dt-\tfrac{1}{2}C\int\int(X_t'-X_s')^2 \times e^{-w|t-s|}dtds+\int f(t)X_t'dt\right], \quad (26)$$

where X' is that function which minimizes the expression [subject for convenience, to $X'(0)=X'(T)=0$ if the time interval is 0 to T]. The variation problem gives the integral equation

$$d^2X'(t)/dt^2=2C\int(X_t'-X_s')e^{-w|t-s|}ds-f(t). \quad (27)$$

Using (27), (26) can be simplified to

$$I=\exp\left[\tfrac{1}{2}\int f(t)X'(t)dt\right]. \quad (28)$$

We need merely solve (27) and substitute into (28). To do this we define

$$Z(t)=\frac{w}{2}\int e^{-w|t-s|}X_s'ds,$$

so that

$$d^2Z(t)/dt^2=w^2[Z(t)-X'(t)],$$

while (27) is

$$d^2X'(t)/dt^2=\frac{4C}{w}[X'(t)-Z(t)]-f(t).$$

The equations are readily separated and solved. The solution for $X'(t)$ substituted into (28) gives, for the case (25),

$$I=\langle\exp[i\mathbf{K}\cdot(\mathbf{X}_\tau-\mathbf{X}_\sigma)]\rangle$$
$$=\exp\left[-\frac{2CK^2}{v^3w}(1-e^{-v|\tau-\sigma|})-\frac{w^2}{2v^2}K^2|\tau-\sigma|\right], \quad (29)$$

where we have made the substitution

$$v^2=w^2+(4C/w). \quad (30)$$

The result is correctly normalized since it is valid for $\mathbf{K}=0$. The integral on K in (22) is a simple Gaussian, so that substitution into A gives

$$A=\pi^{-\frac{1}{2}}\alpha v\int_0^\infty\left[w^2\tau+\frac{v^2-w^2}{v}(1-e^{-v\tau})\right]^{-\frac{1}{2}}e^{-\tau}d\tau. \quad (31)$$

To find B we need $\langle(\mathbf{X}_t-\mathbf{X}_s)^2\rangle$. This can be obtained by expanding both sides of (29) with respect to $\mathbf{K}$ up to order K^2. Therefore

$$\tfrac{1}{3}\langle(\mathbf{X}_\tau-\mathbf{X}_\sigma)^2\rangle=\frac{4C}{v^3w}(1-e^{-|\tau-\sigma|})+\frac{w^2}{v^2}|\tau-\sigma|.$$

The integral in B is now easily performed and the expression simplifies to

$$B=3C/vw. \quad (32)$$

Finally we need E_1, the energy belonging to our action S_1. This is most easily obtained by differentiating both sides of (15) with respect to C. One finds immediately

$$CdE_1/dC=B,$$

so that, in view of (32) and (30), integration gives

$$E_1=\tfrac{3}{2}(v-w),$$

since $E_1=0$ for $C=0$. Since $E_1-B=(3/4v)(v-w)^2$ we obtain finally for our energy expression:

$$E=\frac{3}{4v}(v-w)^2-A, \quad (33)$$

with A given in (31). The quantities v, w can be considered as two parameters which may be varied separately to obtain a minimum.

The integral in A unfortunately cannot be performed in closed form, so that a complete determination of E requires numerical integration. It is, however, possible to obtain approximate expressions in various limiting cases. The case of large α corresponds to large v. The choice $w=0$ leads to an integral

$$A=\pi^{-\frac{1}{2}}\alpha v^{\frac{1}{2}}\int_0^\infty e^{-\tau}d\tau[1-e^{-v\tau}]^{-\frac{1}{2}}=\frac{\alpha\Gamma(1/v)}{v^{\frac{1}{2}}\Gamma(\frac{1}{2}+1/v)}, \quad (34)$$

and $E_1=3v/4$. It corresponds to the use of a fixed harmonic binding potential in (18). For large v, $e^{-v\tau}$ can be neglected, so that $A=\pi^{-\frac{1}{2}}\alpha v^{\frac{1}{2}}$. This corresponds to using a Gaussian trial function in Landau's method. For α less than 5.8 and $w=0$, (33) does not give a minimum unless $v=0$, so that the $w=0$ case does not give a single expression for all ranges of α. In spite of this disadvantage the result with (34) is relatively simple and fairly accurate. For $\alpha>6$, only fairly large v are important, and the asymptotic formula (good to 1 percent for $v>4$),

$$A=\alpha(v/\pi)^{\frac{1}{2}}[1+(2\ln 2)/v],$$

is convenient. Fröhlich, however, considers the discontinuity at $\alpha=6$ as a serious disadvantage, which it is the purpose of this paper to avoid. This we do by choosing w different from zero.

Let us study (33), just for small α, in case w is not zero. The minimum will occur for v near w. Therefore write $v=(1+\epsilon)w$, consider ϵ small, and expand the root in (31). This gives

$$A=\alpha(v/w)\left[1-\epsilon\int_0^\infty\tau^{-\frac{3}{2}}e^{-\tau}(1-e^{-w\tau})d\tau/w\pi^{\frac{1}{2}}+\cdots\right].$$

The integral is

$$2w^{-1}[(1+w)^{\frac{1}{2}}-1]=P. \quad (35)$$

The problem (33) then corresponds, in this order, to minimizing

$$E=\tfrac{3}{4}w\epsilon^2-\alpha-\alpha\epsilon(1-P).$$

That is,

$$\epsilon=2\alpha(1-P)/3w,$$

which is valid for small α only, as ϵ was assumed small. The resulting energy is

$$E=-\alpha-\alpha^2(1-P)^2/3w.$$

Our method therefore gives a correction even for small α. It is least for $w=3$, in which case it gives

$$E=-\alpha-\alpha^2/81=-\alpha-1.23(\alpha/10)^2. \quad (36)$$

It is not sensitive to the choice of w. For example, for $w=1$ the 1.23 falls only to 0.98. The method of Lee and Pines[6] gives exactly this result (36) to this order. The perturbation expansion has been carried to

[6] T. Lee and D. Pines, Phys. Rev. 92, 883 (1953).

second order by Haga[7] who shows that the exact coefficient of the $(\alpha/10)^2$ term should be 1.26, so that our variational method is remarkably accurate for small α.

The opposite extreme of large α corresponds to large v, and, as we shall see, w near 1. Since $v\gg w$ the integral (31) reduces in the first approximation to (34), which we can use in its asymptotic form. The next approximation in w can be obtained by expanding the radical in (31), considering $w/v\ll 1$. Furthermore, $e^{-v\tau}$ is negligible. In this way we get

$$E=\frac{3}{4v}(v-w)^2-\alpha(v/\pi)^{\frac{1}{2}}\left(1+\frac{2\ln 2}{v}-\frac{w^2}{2v}\right). \quad (37)$$

This is minimum, within our approximation of large v, when $w=1$, and $v=(4\alpha^2/9\pi)-(4\ln 2-1)$:

$$E=-\alpha^2/3\pi-3\ln 2-\tfrac{3}{4}=-0.106\alpha^2-2.83. \quad (38)$$

The approximations do not keep E as an upper limit as, unfortunately, the further terms, of order $1/\alpha^2$ are probably positive.

For further numerical work it is probably sufficiently accurate to take $w=1$ for all α, rather than do the extra work needed to minimize this extra variable. This value of w means that the trial S_1 has the same time exponential in the interaction term as does S. For small α, that is, v near 1, the integral can be expanded in a power series in $(v-1)$. The resulting energy is $(w=1)$:

$$E=-\alpha-0.98(\alpha/10)^2-0.60(\alpha/10)^3 -0.14(\alpha/10)^4\cdots. \quad (39)$$

The two expressions (38), (39) fit fairly well near $\alpha=5$. For practical purposes it may suffice to use (39) below $\alpha=5$ and (38) above. If more accuracy than 3 percent is needed near $\alpha=5$ numerical integration of A must be performed. The value of v which gives (39) is

$$v=1+1.14(\alpha/10)+1.35(\alpha/10)^2+1.88(\alpha/10)^3.$$

This may help to choose an appropriate v. For $w=3$ the results are

$$E=-\alpha-1.23(\alpha/10)^2-0.64(\alpha/10)^3\cdots,$$
$$v=3+2.22(\alpha/10)+1.97(\alpha/10)^2\cdots.$$

EFFECTIVE MASS

Another quantity of interest is the effective mass. If the particle moves with a mean group velocity V, its energy should be greater. For small V the energy goes as V^2, and writing it as $mV^2/2$ we call m the effective mass. Since there is an operator analogous to the momentum which commutes with the Hamiltonian, it would be expected that there is a variational principle which minimizes the energy for each momentum. That is, we ought to be able to extend our method to yield an upper limit to the energy for each value of V, or better, of momentum Q. We have not found the expected extension.

If we limit ourselves just to finding the effective mass for low velocities, however, we may proceed in this manner: For a free particle of mass m whose initial coordinate is 0 and final coordinate is $\mathbf{X}_T$ the sum on trajectories is

$$\exp(-mX_T^2/2T). \quad (40)$$

Hence we can study the effective mass for our system by studying the asymptotic form of (6) in the case $\mathbf{X}_T\neq 0$. The asymptotic form should vary for small $\mathbf{X}_T$ as $\exp(-E_0T-mX_T^2/2T)$, its dependence on $\mathbf{X}_T$ determining m. This only requires that (27) be solved for the boundary conditions $\mathbf{X}'=0$ at $t=0$ and $\mathbf{X}'=\mathbf{X}_T$ at $t=T$. There are some confusing complications at the end points so it is easier to proceed as follows. We will put $\mathbf{X}_T=\mathbf{U}T$ so that the propagation (40) is $\exp(-\frac{1}{2}mU^2T)$. [Note that $\mathbf{U}$ is not a physical velocity because t is an artificial parameter in Eq. (5), and is not the time.] That is, we seek the total energy and equate it to $E_0+\frac{1}{2}mU^2$. But if we substitute $\mathbf{X}'=\mathbf{X}''+\mathbf{U}t$ into (27), we see that it is a solution if $\mathbf{X}''$ is. This $\mathbf{X}''$ goes from 0 at $t=0$ to 0 at $t=T$, and is therefore our previous solution. Such a substitution into (26) means that the term involving $\mathbf{f}$ adds a term $\exp(\int t\mathbf{U}\cdot\mathbf{f}dt)$ so that this is the factor by which I is multiplied, aside from normalization. For the $\mathbf{f}$ given in (25) this is $\exp[i\mathbf{K}\cdot\mathbf{U}(\tau-\sigma)]$ so that we now have

$$\langle\exp[i\mathbf{K}\cdot(\mathbf{X}_\tau-\mathbf{X}_\sigma)]\rangle =\exp\left[-\frac{K^2}{2v^2}F(|\tau-\sigma|)+i\mathbf{K}\cdot\mathbf{U}(\tau-\sigma)\right], \quad (41)$$

where

$$F(\tau)=w^2\tau+\frac{v^2-w^2}{v}(1-e^{-v\tau}). \quad (42)$$

Substitution into (22) and (21) gives for A the value

$$A(\mathbf{U})=2^{-\frac{1}{2}}\alpha\int_0^\infty\int(2\pi^2K^2)^{-1}e^{-\tau} \times\exp\left[-\frac{K^2}{2v^2}F(\tau)+i\mathbf{K}\cdot\mathbf{U}\tau\right]d^3\mathbf{K}d\tau. \quad (43)$$

Second differentiation of (41) with respect to $\mathbf{K}$ shows that

$$\langle(\mathbf{X}_t-\mathbf{X}_s)^2\rangle=3F(t-s)v^{-2}+U^2(t-s)^2,$$

so that one obtains for B the value

$$B=\frac{3C}{vw}+\frac{2C}{w^3}U^2.$$

We again find E_1 from $dE_1/dC=B/C$ and $E_1=\frac{1}{2}U^2$ for $C=0$. Thus

$$E_1=\tfrac{3}{2}(v-w)+\tfrac{1}{2}U^2(1+4Cw^{-3}),$$

[7] E. Haga, Progr. Theoret. Phys. (Japan) **11**, 449 (1954).

and our final expression is

$$E=\tfrac{1}{2}U^2+(3/4v)(v-w)^2-A(\mathbf{U}). \tag{44}$$

We next expand $A(\mathbf{U})$ to order U^2 and write the kinetic energy as $mU^2/2$ to find, finally,

$$m=1+\tfrac{1}{3}\pi^{-\frac{1}{2}}\alpha v^3\int_0^\infty [F(\tau)]^{-\frac{3}{2}}e^{-\tau}\tau^2 d\tau. \tag{45}$$

The values of the parameters to use in (45) are those which were previously found to minimize E when $\mathbf{U}=0$.

For small α this gives

$$m=1+\tfrac{1}{6}\alpha+0.025\alpha^2+\cdots \tag{46}$$

for $w=3$, while for $w=1$ the 0.025 becomes 0.023. For large α it becomes

$$m=16\alpha^4/81\pi^4=202(\alpha/10)^4. \tag{47}$$

Our energy values, coming from a minimum principle, are much more accurate than the mass values, whose precision, especially for large α, is hard to judge. Since (46) and (47) do not match well, intermediate values of α require numerical integration of (45).

Lee and Pines[6] have worked with a different type of variational principle. It seems to be nearly as good as ours for α less than about 5, but is poor for larger α (for example, at $\alpha=15$, Lee and Pines find $E_0<-17.6$, while we find $E_0<-26.8$). This appears to contradict their statement that their method is exact for large α. They are referring to a different problem, however, in which the upper momenta are cut off. This means that in S in (7) the function $|\mathbf{X}_t-\mathbf{X}_s|^{-1}$ is replaced by some other function $V(|\mathbf{X}_t-\mathbf{X}_s|)$ which differs for small $|\mathbf{X}_t-\mathbf{X}_s|$. It is evident, for large α, that the best trajectory will be the one that wanders only slightly and the energy will be $2^{-\frac{1}{2}}\alpha V(0)$ in the limit. Their method gives this result in the limit, as ours would also. For the case where V is singular, so $V(0)$ does not exist their method is not exact, and it is inaccurate for for intermediate values of α even if $V(0)$ exists, if V has steep walls.

The method is readily extended to cases in which the photon frequencies are not constant, and the coupling is not just proportional to K^{-1}. The same trial action S_1 can be used, but the integral for A becomes more complicated. For the Hamiltonian

$$H=\tfrac{1}{2}P^2+\textstyle\sum_K \omega_K a_K^+ a_K+V^{-\frac{1}{2}}\sum_K[C_K^* a_K^+\exp(-i\mathbf{K}\cdot\mathbf{X}) \\ +C_K a_K\exp(+i\mathbf{K}\cdot\mathbf{X})],$$

Eq. (33) still holds; the only change is that the integral for A becomes

$$A=\int\int_0^\infty \exp\left[-\omega_K\tau-\frac{K^2}{2v^2}F(\tau)\right]|C_K|^2 d\tau d^3\mathbf{K}(2\pi)^{-3},$$

where $F(\tau)$ is given in (42).

An attempt has been made to apply this method to meson problems. The case of scalar nucleons interacting by scalar mesons seems tractable, but the greater complexity of the more realistic problems shows the need for further development.

We are limited in our choice of S_1 to quadratic functionals, for those are the only ones we can evaluate directly as path integrals. It would be desirable to find out how this method may be expressed in conventional notation, for a wider class of trial functionals might thereby become available.

I am indebted to H. Fröhlich for bringing the problem to my attention, and for his comments on it, and to G. Speisman for emphasizing the importance of the general inequality (10).

Note added in proof.—Professor Frölich and Professor Pines have kindly informed me that S. I. Pekar [Zhur. Eksptl. i Teort. Fiz. **19**, 796 (1949)] has calculated the limiting values of energy and mass for large α, by an adiabatic approximation. The energy is $-0.1088\alpha^2$ and the mass is $232(\alpha/10)^4$. Therefore our variational method gives an error of only 3 percent in the energy and 15 percent in the mass for large α, and presumably smaller errors for smaller α.

Vol. XXIV (1963) ACTA PHYSICA POLONICA Fasc. 6 (12)

QUANTUM THEORY OF GRAVITATION*

By R. P. Feynman

(Received July 3, 1963)

My subject is the quantum theory of gravitation. My interest in it is primarily in the relation of one part of nature to another. There's a certain irrationality to any work in gravitation, so it's hard to explain why you do any of it; for example, as far as quantum effects are concerned let us consider the effect of the gravitational attraction between an electron and a proton in a hydrogen atom; it changes the energy a little bit. Changing the energy of a quantum system means that the phase of the wave function is slowly shifted relative to what it would have been were no perturbation present. The effect of gravitation on the hydrogen atom is to shift the phase by 43 seconds of phase in every hundred times the lifetime of the universe! An atom made purely by gravitation, let us say two neutrons held together by gravitation, has a Bohr orbit of 10^8 light years. The energy of this system is 10^{-70} rydbergs. I wish to discuss here the possibility of calculating the Lamb correction to this thing, an energy, of the order 10^{-120}. This irrationality is shown also in the strange gadgets of Prof. Weber, in the absurd creations of Prof. Wheeler and other such things, because the dimensions are so peculiar. It is therefore clear that the problem we are working on is not the correct problem; the correct problem is what determines the size of gravitation? But since I am among equally irrational men I won't be criticized I hope for the fact that there is no possible, practical reason for making these calculations.

I am limiting myself to not discussing the questions of quantum geometry nor what happens when the fields are of very short wave length. I am not trying to discuss any problems which we don't already have in present quantum field theory of other fields, not that I believe that gravitation is incapable of solving the problems that we have in the present theory, but because I wish to limit my subject. I suppose that no wave lengths are shorter than one-millionth of the Compton wave length of a proton, and therefore it is legitimate to analyze everything in perturbation approximation; and I will carry out the perturbation approximation as far as I can in every direction, so that we can have as many terms as we want, which means that we can go to ten to the minus two-hundred and something rydbergs.

I am investigating this subject despite the real difficulty that there are no experiments. Therefore there is so real challenge to compute true, physical situations. And so I made

* Based on a tape-recording of Professor Feynman's lecture at the *Conference on Relativistic Theories of Gravitation*, Jabłonna, July, 1962. — Ed.

believe that there were experiments; I imagined that there were a lot of experiments and that the gravitational constant was more like the electrical constant and that they were coming up with data on the various gravitating atoms, and so forth; and that it was a challenge to calculate whether the theory agreed with the data. So that in each case I gave myself a specific physical problem; not a question, what happens in a quantized geometry, how do you define an energy tensor *etc.*, unless that question was necessary to the solution of the physical problem, so please appreciate that the plan of the attack is a succession of increasingly complex physical problems; if I could do one, then I was finished, and I went to a harder one imagining the experimenters were getting into more and more complicated situations. Also I decided not to investigate what I would call familiar difficulties. The quantum electrodynamics diverges; if this theory diverges, it's not something to be investigated unless it produces any specific difficulties associated with gravitation. In short, I was looking entirely for unfamiliar (that is, unfamiliar to meson physics) difficulties. For example, it's immediately remarked that the theory is non-linear. This is not at all an unfamiliar difficulty; the theory, for example, of the spin 1/2 particles interacting with the electromagnetic field has a coupling term $\bar{\psi}A\psi$ which involves three fields and is therefore non-linear; that's not a new thing at all. Now, I thought that this would be very easy and I'd just go ahead and do it, and here's what I planned. I started with the Lagrangian of Einstein for the interacting field of gravity and I had to make some definition for the matter since I'm dealing with real bodies and make up my mind what the matter was made of; and then later I would check whether the results that I have depend on the specific choice or they are more powerful. I can only do one example at a time; I took spin zero matter; then, since I'm going to make a perturbation theory, just as we do in quantum electrodynamics, where it is allowed (it is especially more allowed in gravity where the coupling constant is smaller), $g_{\mu\nu}$ is written as flat space as if there were no gravity plus $\varkappa$ times $h_{\mu\nu}$, where $\varkappa$ is the square root of the gravitational constant. Then, if this is substituted in the Lagrangian, one gets a big mess, which is outlined here.

$$\mathcal{L} = \frac{1}{\varkappa^2}\int R\sqrt{g}\,d\tau + \frac{1}{2}\int\left(\sqrt{g}\,g^{\mu\nu}\varphi_{,\mu}\varphi_{,\nu} - m^2\sqrt{g}\,\varphi^2\right)d\tau \tag{1}$$

$$g_{\mu\nu} = \delta_{\mu\nu} + \varkappa h_{\mu\nu}.$$

Substituting and expanding, and simplifying the results by a notation (a bar over a tensor means

$$\bar{x}_{\mu\nu} \equiv \frac{1}{2}\left(x_{\mu\nu} + x_{\nu\mu} - \delta_{\mu\nu}x_{\sigma\sigma}\right);$$

notice that if $x_{\mu\nu}$ is symmetric, $\bar{\bar{x}}_{\mu\nu} = x_{\mu\nu}$) we get

$$\mathcal{L} = \int\left(h_{\mu\nu,\sigma}\bar{h}_{\mu\nu,\sigma} - 2\bar{h}_{\mu\sigma,\sigma}\bar{h}_{\mu\sigma,\sigma}\right) + \frac{1}{2}\int\left(\varphi_{,\mu}^2 - m^2\varphi^2\right)d\tau +$$

$$+ \varkappa\int\left(\bar{h}_{\mu\nu}\varphi_{,\mu}\varphi_{,\nu} - m^2\,\frac{1}{2}\,h_{\sigma\sigma}\varphi^2\right) + \varkappa\int\text{“}hhh\text{”} + \varkappa^2\int\text{“}hh\varphi\varphi\text{”} + \text{etc.} \tag{2}$$

First, there are terms which are quadratic in h; then there are terms which are quadratic in φ, the spin zero meson field variable; then there are terms which are more complicated than quadratic; for example, here is a term with two φ's and one h, which I will write $h\varphi\varphi$ (I have written that one out, in particular); there are terms with three h's; then there are terms which involve two h's and two φ's; and so on and so on with more and moie complicated terms. The first two terms are considered as the free Lagrangian of the gravitational field and of the matter.

Now we look first at what we would want to solve problem classically, we take the variation of this with respect to h, from the first term we produce a certain combination of second derivatives, and on the other side a mess involving higher orders than first. And the same with the φ, of course.

$$h_{\mu\nu,\sigma\sigma}-\bar{h}_{\sigma\nu,\sigma\mu}-\bar{h}_{\sigma\mu,\sigma\nu} = \bar{S}_{\mu\nu}(h, \varphi) \tag{3}$$

$$\varphi_{,\sigma\sigma}-m^2\varphi = \chi(\varphi, h). \tag{4}$$

We will speak in the following way: (3) is a wave equation, of which $S_{\mu\nu}$ is the source, just like (4) is the wave equation of which χ is the source. The problem is to solve those equations in succession, and to use the usual methods of calculation of the quantum theory. Inasmuch as I wanted to get into the minimum of difficulties, I just took a guess that I use the same plan as I do in electricity; and the plan in electricity leads to the following suggestion here: that if you have a source, you divide by the operator on the left side of (3) in momentum space to get the propagator field. So I have to solve this equation (3). But as you all know it is singular; the entire Lagrangian in the beginning was invariant under a complicated transformation of g, which in the form of h is the following; if you add to h a gradient plus more, the entire system is invariant:

$$h'_{\mu\nu} = h_{\mu\nu}+2\xi_{\mu,\nu}+2h_{\mu\sigma}\,\xi_{\sigma,\nu}+\xi_{\sigma}\,h_{\mu\nu,\sigma}, \tag{5}$$

where ξ_μ is arbitrary, and μ and ν should be made symmetric in all these equations. As a consequence of this same invariance in the complet Lagrangian one can show that the source $S_{\mu\nu}$ must have zero divergence $S_{\mu\nu,\nu} = 0$. In fact equations (3) would not be consistent without this condition as can be seen by barring both sides and taking the divergence — the left side vanishes identically. Now, because of the invariance of the equations, in the same way that the Maxwell equations cannot be solved to get a unique vector potential — so these can't be solved and we can't get a unique propagator. But because of the invariance under the transformation some arbitrary choice of a condition on $h_{\mu\nu}$ can be made, analogous to the Lorentz condition $A_{\mu,\mu}=0$ in quantum electrodynamics. Making the simplest choice which I know, I make choice $\bar{h}_{\mu\sigma,\sigma} = 0$. This is four conditions and I have free the four variables ξ_μ that I can adjust to make the condition satisfied by $h'_{\mu\nu}$. Then this equation (3) is very simple, because two terms in (3) fall away and all we have is that the d'Alemberian of h is equal to S. Therefore the generating field from a source $S_{\mu\nu}$ will equal the $\bar{S}_{\mu\nu}$ times $1/k^2$ in Fourier series, where k^2 is the square of the frequency, wave vector; the time part might be called the frequency ω, the space part $\boldsymbol{k}$. This is the analogue of the equation in electricity that says that the field is $1/k^2$ times the current. In the method of quantum field theory, you have a source which generates something, and that may interact later with something else; the iteraction, of course, is $S_{\mu\nu}\, h_{\mu\nu}$; so that, I say, one source may create a potential which acts on another source. So, to take the very simplest example of two interacting systems, let's say S and S', the result would be the following: h would be generated by $S_{\mu\nu}$, and then it would interact with $S'_{\mu\nu}$, so we would get for the interaction of two systems, of two particles, the fundamental interaction that we investigate

$$\varkappa^2 \bar{S}_{\mu\nu} \frac{1}{k^2} S'_{\mu\nu}. \tag{6}$$

This represents the law of gravitational interaction expressed by means of an interchange of a virtual graviton. To understand the theory better and to see how far we already arrived we expand it out in components. Let index 4 represent the time, and 3 the direction of $\mathbf{k}$, so that 1 and 2 are transverse. The condition $k_\mu S_{\mu\nu} = 0$ becomes $\omega S_{4\nu} = k S_{3\nu}$ where k is the magnitude of $\mathbf{k}$. Using this, many of the terms involving number 3 component of S can be replaced by terms in number 4 components. After some rearranging there results

$$-2\bar{S}_{\mu\nu} \frac{1}{k^2} S_{\mu\nu} = \frac{1}{\mathrm{k}^2}\,[S_{44} S'_{44}] + \frac{1}{\mathrm{k}^2}\,[S_{44}(S'_{11}+S'_{22}) + S'_{44}(S_{11}+S_{22}) +$$

$$+ S_{43} S'_{43} - 4S_{41} S'_{41} - 4S_{42} S'_{42}] + \frac{1}{\mathrm{k}^2 - \omega + i\varepsilon}\,[(S_{11}-S_{22})\,(S'_{11}-S'_{22}) + 4S_{12}S'_{12}]. \tag{7}$$

There is a singular point in the last term when $\omega = \mathrm{k}$, and to be precise we put in the $+i\varepsilon$ as is well-known from electrodynamics. You note that in the first two terms instead of one over a four-dimensional $\omega^2 - \mathbf{k}^2$ we have here just $1/\mathbf{k}^2$, the momentum itself. S_{44} is the energy density, so this first term represents the two energy densities interacting with no ω dependence which means, in the Fourier transform an interaction instantaneous in time; and $1/\mathbf{k}^2$ means $1/r$ in space, so there's an instantaneous $1/r$ interaction between masses, Newton's law. In the next term there's another instantaneous term which says that Newton's mass law should be corrected by some other components analogous to a kind of magnetic interaction (not quite analogous because the magnetic interaction in electricity already involves a $\mathrm{k}^2 - \omega^2 + i\varepsilon$ propagator rather than just k^2. But the $\mathrm{k}^2 - \omega^2 + it$ in gravitation comes even later and is a much smaller term which involves velocities to the fourth). So if we really wanted to do problems with atoms that were held together gravitationally it would be very easy; we would take the first term, and possibly even the second as the interaction. Being instantaneous, it can be put directly into a Schrödinger equation, analogous to the e^2/r term for electrical interaction. And that take care of gravitation to a very high accuracy, without a quantized field theory at all. However, for still higher accuracy we have to do the radiative corrections, which come from the last term.

Radiation of free gravitons corresponds to the situation that there is a pole in the propagator. There is a pole in the last term when $\omega = \mathrm{k}$, of course, which means that the wave number and the frequency are related as for a mass zero particle. The residue of the pole, we see, is the product of two terms; which means that there are two kinds of waves, one generated by $S_{11} - S_{22}$ and the other generated by S_{12}, and so we have two kinds of transverse polarized waves, that is there are two polarization states for the graviton. The linear combination $S_{11} - S_{22} \pm 2iS_{12}$ vary with angle Θ of rotation in the $1-2$ plane as $e^{\pm 2i\theta}$ so the gravitaton has spin 2, component ± 2 along direction of polarization. Everything is clear directly from the expression (7); I just wanted to illustrate that the propagator (6) of quantum mechanics and all that we know about the classical situation are in evident coincidence.

In order to proceed to make specific calculations by means of diagrams, beside the propagator we need to know just what the junctions are, in other words just what the S's are for a particular problem; and I shall just illustrate how that's done in one example. It is done by looking at the non-quadratic terms in the Lagrangian I've written one out completely. This one has an h and two φ's in the Lagrangian (2). The rules of the quantum mechanics for writting this thing are to look at the h and two φ's: one φ each refers to the in and out particle, and the one h corresponds to the graviton; so we immediately see in that term a two particle interaction through a graviton (see Fig. 1). And we can immediately

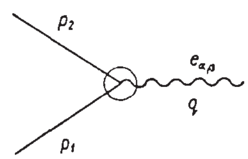


Fig. 1

read off the answer for the interaction this way: if the p_1 and p_2 are the momenta of the particles and q the momentum of the graviton; and $e_{\alpha\beta}$ is the polarization tensor of the plane wave representing the graviton, that is $h_{\alpha\beta} = e_{\alpha\beta}\, e^{iq\cdot x}$, the Fourier expansion of this term gives the amplitude for the coupling of two particles to a graviton

$$p^1_\mu p^2_\nu \bar{e}_{\mu\nu} - \frac{1}{2} m^2 e_{\sigma\sigma}. \tag{8}$$

So this is a coupling of matter to gravity; it is first order, and then there are higher terms; but the point I'm trying to make is that there is no mystery about what to write down — everything is perfectly clear, from the Lagrangian. We have the propagator, we have the couplings, we can write everything. A term like hhh implies a definite formula for the interaction of three gravitons; it is very complicated, and I won't write it down, but you can read it right off directly by substituting momenta for the gradients. That such a term exists is, of course, natural, because gravity interacts with any kind of energy, including its own, so if it interacts with an object-particles it will interact with gravitons; so this is the scattering of a graviton in a gravitational field, which must exist. So that everything is directly readable and all we have to do is proceed to find out if we get a sensible physics. I've already indicated that the physics of direct interactions is sensible; and I go ahead now to compute a number of other things.

To take just one example, we compute the Compton effect, or the analogue rather, of the Compton effect, in which a graviton comes in and out on a particle. The amplitude for this is a sum of terms corresponding to the diagrams of Fig. 2. The amplidute for the first diagram of Fig. 2 is the coupling (8) times the propagator for the intermediate meson which reads $(p^2-m^2)^{-1}$, which is the Fourier transform of the equation (4) which is the propagation of the spin zero paiticle. Then there is another coupling of the same form as (8). We multiply these together, to get the amplitude for that diagram

$$\left(p^2_\mu p_\nu \bar{e}^b_{\mu\nu} - \frac{1}{2}\, e_{\mu\mu}\, m^2\right) \frac{1}{p^2-m^2} \left(p_\sigma\, p^1_\tau\, e^a_{\sigma\tau} - \frac{1}{2}\, \bar{e}_{\sigma\sigma} m^2\right),$$

where we should substitute $p = p^2+q^b = p^1+q^a$. Then you must add similar contributions from the other diagrams.

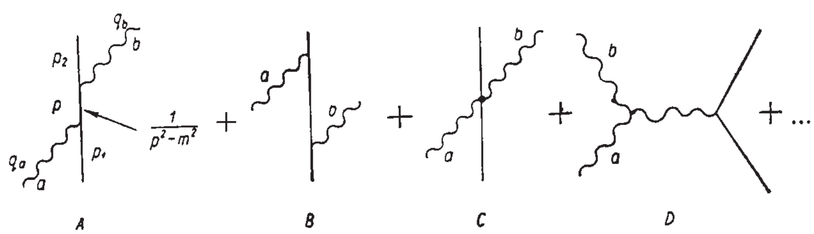


Fig. 2

The third one comes in because there are terms with two h's and two φ's in the Lagrangian. One adds the four diagrams together and gets an answer for the Compton effect. It is rather simple, and quite interesting; that it is simple is what is interesting, because the labour is fa ntastic in all these things.

But the thing I would like to emphasize is this; in this problem we used a certain wave $e^a_{\alpha\beta}$ for the incoming graviton number "a" say; the question is could we use a different one? According to the theory, it should really be invariant under coordinate transformations and so on, but what it corresponds to here is the analogue of gauge invariance, that you can add to the potential a gradient (see (5)). And therefore it should be that if I changed $e_{\alpha\beta}$ of a particular graviton to $e_{\alpha\beta}+q_\alpha\xi_\beta$ where ξ is arbitrary, and q_α is the momentum of the graviton, there should be no change in the physics. In short, the amplitude should be unchanged; and it is. The amplitude for this particular process is what I call gauge-invariant, or coordinate-transforming invariant. At first sight this is somewhat puzzling, because you would have expected that the invariance law of the whole thing is more complicated, including the last two terms in (5), which I seem to have omitted. But those terms have been included; you see asymptotically all you have to do is worry about the second term, the last two in h's times ξ's are in fact generated by the last diagram, Fig. 2D; when I put a gradient in here for this one, what this means is if I put for the incoming wave a pure gradient, I should get zero. If I put the gradient $q_\alpha\xi_\beta$ in for $e^a_{\alpha\beta}$ on this term D, I get a coupling between ξ and the other field $e^a_{\alpha\beta}$ because of the three graviton coupling. The result, as far as the matter line is concerned is that it is acted on in first order by a resultant field $e^b_{\mu\sigma}\,\xi_\sigma\, q^a_\nu +$ $+\ \frac{1}{2}\, q^b_\sigma\, e_{\mu\nu}\, \xi_\sigma$ which is just the last two terms in (5). The rule is that the field which acts on the matter itself must be invariant the way described by (5); but here in Fig. 2 I've already calculated all the corrections, the generator and all the necessary non-linear modifications if I take all the diagrams into account. In short, asymptotically far away if I include all kinds of diagrams such as D, the invariance need be checked only for a pure gradient added to an incoming wave. It takes care of the non-linearities by calculating them through the interaction.

I would like, now, to emphasize one more point that is very important for our later discussion. If I add a gradient, I said, the result was zero. Let's call a the one graviton coming in and b the other one in every diagram. The result is zero if I use a gradient for a, only if b is a free graviton with no source; that is if it is either really an honest graviton with $(q^b)^2 = 0$, or a pure potential, which is a solution of the free wave equation. That is unlike electrodynamics, where the field b could have been any potential at all and adding a gradient to a would have made no difference. But in gravity, it must be that b is a pure wave; the reason is very simple. There is no way to avoid this by changing any propagators; this is not a disease — there is a physical reason. The reason can be seen as follows: If this b had a source let me modify my diagrams to show the source of b, suppose some other matter particle made the b, so we add onto each b line a matter line at the end, like Fig. 3a. (*E.g.* Fig. 2a becomes Fig. 3b *etc.*)

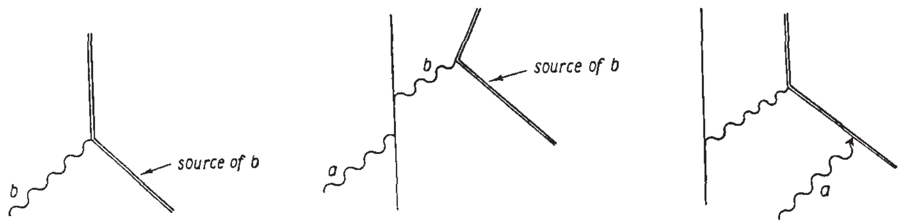


Fig. 3

Now, if b isn't a free wave, but it had a source, the situation is this. If this "a" field is taken as a gradient field which operates everywhere on everything in the diagram it should give zero. But we forgot something; there's another type of diagram, if the "a" is supposed to act on everything, one of which looks like Fig. 3c, in which the "a" itself acts on the source of b and then b comes over to interact with the original matter. In other words, among all the diagrams where there is a source, there's also these of type 3c. The sum of all diagrams is zero; but the sum of those like Fig. 2 without those of type 3c is not zero, and therefore if I were to just calculate the diagrams of Fig. 2 and forget about the source of b and then put a gradient in for "a" the result cannot be zero, but must be getting ready to cancel the terms from the likes of 3c when I do it right. That will turn out to be an important point to emphasize. I have done a lot of problems like this, without closed loops but I won't bore you with all the problems and answers; there's nothing new, I mean nothing interesting, in the sense that no apparent difficulties arise.

However, the next step is to take situations in which we have what we call closed loops, or rings, or circuits, in which not all momenta of the problem are defined. Let me just mention something. I've analyzed this method both by doing a number of problems, and by a mathematical high-class elegant technique — I can do high class mathematics too, but I don't believe in it, that's the difference. I have to check it in a problem. I can prove that no matter how complicated the problem is, if you take it in the order in which there are no rings, in which every momentum is determined, the invariance is satisfied, the system is independent of what choice I made of gauge and of the propagator I made in the beginning; and everything is all right, there are no difficulties. I emphasize that this contains all the classical cases, and so I 'm really saying there are no difficulties in the classical gravitation theory. This is not meant as a grand discovery, because after all, you've been worrying about all these difficulties that I say don't exist, but only for you to get an idea of the calibration — what I mean by difficulties! If we take the next case, let's say the interaction of two particles in a higher order, then you get diagrams of which I'll only begin to write a few of them. One that looks like this in which two gravitons are exchanged,

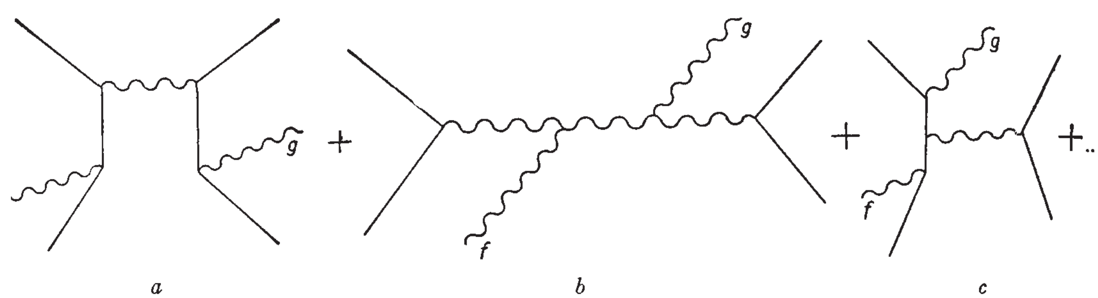


Fig. 4

or, for instance, a graviton gets split into two gravitons and then come back — these are only the beginning of a whole series of frightening-looking pictures, which correspond to the problem of calculating the Lamb shift, or the radiative corrections to the hydrogen atom. When I tried to do this, I did it in a straightforward way, following all the rules, putting in the propagator $1/k^2$, and so on. I had some difficulties, the thing didn't look gauge invariant but that had to do with the way I was making the cutoffs, because the stuff is infinite. Shortage of time does'nt permit me to explain the way I got around all those things, because in spite of getting around all those things the result is nevertheless definitely incorrect. It's gauge-invariant, it's perfectly O.K. looking, but it is definitely incorrect. The reason I knew it was incorrect is the following. In order to get it gauge-invariant, I had to do a lot of pushing and pulling, and I got the feeling that the thing might not be unique. I figured that maybe somebody else could do it another way or something, and I was rather suspicious, so I tried to get more tests for it; and a student of mine, by the name of Yura, tested to see if it was unitary; and what that means is the following: Let me take instead of this scattering problem, a problem of Fig. 4 in which time runs vertically, a problem which gives the same diagrams but in which time is running horizontally, which is the annihilation of a pair, to produce another pair, and we are calculating second order corrections to that problem. Let's suppose for simplicity that in the final state the pair is in the same state as before.

Then, adding all these diagrams gives the amplitude that if you have a pair, particle and antiparticle, they annihilate and recreate themselves; in other words it's the amplitude that the pair is still in the same state as a function of time. The amplitude to remain in the same state for a time T in general is of the form

$$e^{-i\left(E_0-i\frac{\gamma}{2}\right)T}$$

you see that the imaginary part of the phase goes as $e^{-\frac{\gamma}{2}T}$; which means that the probability of being in a state must decrease with time. Why does the probability decrease in time? Because there's another possibility, namely, these two objects could come together, annihilate, and produce a real pair of gravitons. Therefore, it is necessary that this decay rate of the closed loop diagrams in Fig. 4 that I obtain by directly finding the imaginary part of the sum agrees with another thing I can calculate independently, without looking at the closed loop diagrams. Namely, what is the rate at which a particle and antiparticle annihilate into two gravitons? And this is very easy to calculate (same set of diagrams as Fig. 2, only turned on its side). I calculated this rate from Fig. 2, checked whether this rate agrees with the rate at which the probability of the two particles staying the same decreases (imaginary part of Fig. 4), and it does not check. Somethin'gs the matter.

This made me investigate the entire subject in great detail to find out what the trouble is. I discovered in the process two things. First, I discovered a number of theorems, which as far as I know are new, which relate closed loop diagrams and diagrams without closed loop diagrams (I shall call the latter diagrams "trees"). The unitarity relation which I have just been describing, is one connection between a closed loop diagram and a tree; but I found a whole lot of other ones, and this gives me more tests on my machinery. So let me just tell you a little bit about this theorem, which gives other rules. It is rather interesting. As a matter of fact, I proved that if you have a diagram with rings in it there are enough theorems altogether, so that you can express any diagram with circuits completely in terms of diagrams with trees and with all momenta for tree diagrams in physically attainable regions and on the mass shell. The demonstration is remarkably easy. There are several ways of demonstrating it; I'll only chose one. Things propagate from one place to another, as I said, with amplitude $1/k^2$. When translated into space, that's a certain propagation function which you might call $K_+(1, 2)$, a function of two positions, 1, 2, in space-time. It represents, in the past, incoming waves and in the future, it represents outgoing waves; so you have waves come in and out; and that's the conventional propagator, with the $i\varepsilon$ and so on, as usually represented. However, this is only a solution of the propagators's equation, the wave equation I mean; it is a special solution, as you all know. There are other solutions; for instance there is a solution which is purely retarded, which I'll call K_{ret} and which exists only inside the future light-cone. Now, if you have two Green's functions for the same equation they must differ by some solution of the homogeneous equation, say K_x. That means K_x is a solution of the free wave equation and $K_+ = K_{\text{ret}} + K_x$. In a ring like Fig. 4a we have a whole product of these K_+'s. For example, for four points 1, 2, 3, 4 in a ring we have a product like this: $K_+(1, 2)K_+(2, 3)K_+(3, 4)K_+(4, 1)$ (all K's are not the same, some of them belong to the gravitons and some are propagators for the particles and so on).

But now let us see what happens if we were to replace one (or more) of these K_+ by K_x, say $K_+(1, 2)$ is $K_x(1, 2)$? Then between 1, 2 we have just free particles, you've broken the ring; you've got an open diagram, because K_x is free wave solution, and this means it's an integral over all real momenta of free particles, on the mass shell and perfectly honest. Therefore if we replace one of K_+ by K_x then that particular line is opened; and the process is changed to one in which there is a forward scattering of an extra particle; there's a fake particle that belongs to this propagator that has to be integrated over, but it's a free diagram — it is now a tree, and therefore perfectly definite and unique to calculate. But I said that I could open every diagram; the reason is this. First I note that if I put K_{ret} for every K in a ring, I get zero

$$K_{ret}(1, 2)K_{ret}(2, 3)K_{ret}(3, 4)K_{ret}(4, 1) = 0 \tag{9}$$

for to be non zero t_1 must be greater than t_2, $t_2 > t_3$, $t_3 > t_4$ and $t_4 > t_1$ which is impossible. Now make the substitution $K_{ret} = K_+ - K_x$ in (9). You get either all K_+ in each factor, which is the closed loop we want; or at least one K_x, which are represented by tree diagrams. Since the sum is zero, closed loops can be represented as integrals over tree diagrams. I was surprised I had never noticed this thing before.

Well, then I checked whether these diagrams of Fig. 4 when opened into trees agreed with the theorem. I mean I hoped that the theorem proved for other meson theories would agree in principle for the gravity case, such that on opening a virtual graviton line the tree would correspond to forward scattering of free graviton waves. And it does not work in the gravity case. But, you say, how could it fail, after you just demonstrated that it ought to work? The reason it fails is the following: This argument has to do with the position of the poles in the propagators; a typical propagator is a factor $1/(k^2-m^2+i\varepsilon)$, the $+i\varepsilon$ due to the poles, and all I'm doing here is changing the rule about the poles and picking up an extra delta function $\delta(k^2-m^2)$ as a consequence, which is the free wave coming in and out. What I want these free waves to represent in the gravity case are physical gravitons and not something wrong. They do represent waves of $q^2 = 0$ of course, but, as it turns out, not with the correct polarization to be free gravitons. I'd like to show it. It has to do with the numerator, not the denominator. You see the propagator that I wrote before, which was $S_{\mu\nu}$ times $1/(k^2+i\varepsilon)$ times $\bar{S}'_{\mu\nu}$, is being replaced by $S_{\mu\nu}\,\delta(q^2)\bar{S}'_{\mu\nu}$. Now when I make $q^2 = 0$ I have a free wave instead of arbitrary momentum. This should be a real graviton or else there's going to be physical trouble. It ins't; although it is of zero momentum, it is not transverse. It does not make any difference in understanding the point so forget one index in $S_{\mu\nu}$ — it's a lot of extra work to carry the other index so just imagine there's one index: $S_\mu S_\mu \delta(q^2)$. This combination $S_\mu S'_\mu$, is $S_4S'_4 - S_3S'_3 - S_1S'_1 - S_2S'_2$, where 4 is the time and 3 is the direction, say, of momentum of the four-vector q. Then 1 and 2 are transverse, and those are the only two we want. (Please appreciate I removed one index — I can make it more elaborate, but it is the same idea.) That is we want only $-S_1S'_1 - S_2S'_2$ instead of the sum over four. Now what about this extra term $S_4S'_4 - S_3S'_3$? Well, it is $S_4 - S_3$ times $S'_4 + S'_3$ plus $S_4 + S_3$ times $S'_4 - S'_3$. But $S_4 - S_3$ is proprtional to $q_\mu S_\mu$ (suppressing one index) because q_4 in this notation is the frequency and equals q_3, if we assume the 3-direction is the direction of the momentum. So $S_4 - S_3$ is the response of the system to a gradient potential, which we proved was zero in our invariance discussion. Therefore, we have shown $(S_4-S_3)/(S_4'+S_3') = 0$ and this should be accounted for by purely transverse wave contributions. But it ins't, and it isn't because *the proof that the response to a gradient potential is zero required that the other particle that was interacting was an honest free graviton.* And four plus three in $S_4'+S_3'$ is not honest — it's not transverse, it is not a correct kind of graviton. You see, the only way you can get a polarization 4+5 going in the 4—3 direction is to have what I call longitudinal response; it's not a transverse wave. Such a wave could only be generated by an artificial source here of some silly kind; it is not a free wave. When there's an artificial source for one graviton, even the another is a pure gradient, the sum of all the diagrams does not give zero. If the beam is not exactly that of a free wave, perfectly transverse and everything, the argument that the gradient has to be zero must fail, for the reason outlined previously.

Although this gradient for S_4-S_3 is what I want and I hoped it was going to be zero I forgot that the other end of it — $S_4'+S_3'$ is a funny wave which is not a gradient, and which is not a free wave — and therefore you do not get zero and should not get zero, and something is fundamentally wrong.

Incidentally I investigated further and discovered another very interesting point. There is another theory, more well-known to meson physicists, called the Yang-Mills theory, and I take the one with zero mass; it is a special theory that has never been investigated in great detail. It is very analogous to gravitation; instead of the coordinate transformation group being the source of everything, it's the isotopic spin rotation group that's the source of everything. It is a non-linear theory, that's like the gravitation theory, and so forth. At the suggestion of Gell-Mann I looked at the theory of Yang-Mills with zero mass, which has a kind of gauge group and everything the same; and found exactly the same difficulty. And therefore in meson theory it was not strictly unknown difficulty, because it should have been noticed by meson physicists who had been fooling around the Yang-Mills theory. They had not noticed it because they're practical, and the Yang-Mills theory with zero mass obviously does not exist, because a zero mass field would be obvious; it would come out of nuclei right away. So they didn't take the case of zero mass and investigate it carefully. But this disease which I discovered here is a disease which exist in other theories. So at least there is one good thing: gravity isn't alone in this difficulty. This observation that Yang-Mills was also in trouble was of very great advantage to me; it made everything much easier in trying to straighten out the troubles of the preceding paragraph, for several reasons. The main reason is if you have two examples of the same disease, then there are many things you don't worry about. You see, if there is something different in the two theories it is not caused by that. For example, for gravity, in front of the second derivatives of $g_{\mu\nu}$ in the Lagrangian there are other g's, the field itself. I kept worrying something was going to happen from that. In the Yang-Mills theory this is not so, that's not the cause of the trouble, and so on. That's one advantage — it limits the number of possibilities. And the second great advantage was that the Yang-Mills theory is enormously easier to compute with than the gravity theory, and therefore I continued most of my investigations on the Yang-Mills theory, with the idea, if I ever cure that one, I'll turn around and cure the other. Because I can demonstrate one thing; line for line it's a translation like music transcribed to a different score; everything has its analogue precisely, so it is a very good example to work with. Incidentally, to give you some idea of the difference in order to calculate this diagram Fig. 4b the Yang-Mills case took me about a day; to calculate the diagram in the case of gravitation I tried again and again and was never able to do it; and it was finally put on a computing machine—I don't mean the arithmetic, I mean the algebra of all the terms coming in, just the algebra; I did the integrals myself later, but the algebra of the thing was done on a machine by John Matthews, so I couldn't have done it by hand. In fact, I think it's historically interesting that it's the first problem in algebra that I know of that was done on a machine that has not been done by hand.

Well, what then, now you have the difficulty; how do you cure it? Well I tried the following idea: I assumed the tree theorem to be true, and used it in reverse. If every closed ring diagram can be expressed as trees, and if trees produce no trouble and can be computed, then all you have to do is to say that the closed loop diagram is the sum of the corresponding tree diagrams, that it should be. Finally in each tree diagram for which a graviton line has been opened, take only real transverse graviton to represent that term. This then serves as the definition of how to calculate closed-loop diagrams; the old rules, involving a propagator $1/k^2+i\varepsilon$ *etc.* being superseded. The advantage of this is, first, that it will be gauge invariant, second, it will be unitary, because unitarity is a relation between a closed diagram and an open one, and is one of the class of relations I was talking about, so there's no difficulty. And third, it's completely unique as to what the answer is; there's no arbitrary fiddling around with different gauges and so forth, in the inside ring as there was before. So that's the plan.

Now, the plan requires, however, one more point. It's true that we proved here that every ring diagram can be broken up into a whole lot of trees; but, a given tree is not gauge invariant. For instance the tree diagram of Fig. 2A is not. Each one of the four diagrams of Fig. 2 is not gauge-invariant, nor is any combination of them except the sum of all four. So the thing is the following. Suppose I take all the processes, all of them that belong together in a given order; for example, all the diagrams of fourth order, of which Fig. 4 illustrates three; I break the whole mess into trees, lots of trees. Then I must gather

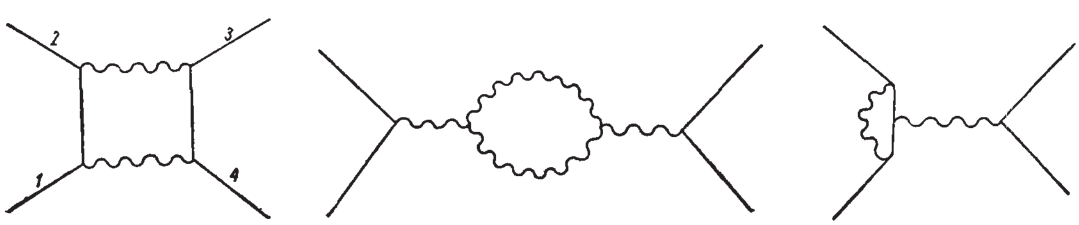


Fig. 5

the trees into baskets again, so that each basket contains the total of all of the diagrams of some specific process (for example the four diagrams of Fig. 2), you see, not just some particular tree diagram but the complet set for some process. The business of gathering the tree diagrams together in bunches representing all diagrams for complet processes is important, for only such a complet set is gauge invariant. The question is: Will any odd tree diagrams be left out or can they all be gathered into processes? The question is: Can we express the closed ring diagrams for some process into a sum over various other processes of tree diagrams for these processes?

Well, in the case with one ring only, I am sure it can be done, I proved it can be done and I have done it and it's all fine. And therefore the problem with one ring is fundamentally solved; because we say, you express it in terms of open parts, you find the processes that they correspond to, compute each process and add them together.

You might be interested in what the rule is for one ring; it's the sum of several pieces: first it is the sum of all the processes which you get in the lower order, in which you scatter one extra particle from the system. For instance, in Fig. 4 we have the rings for two particles scattering. There is no external graviton but there are two internal ones; now we compute in the same order a new problem in which there are two particles scattering, but while that's happening another particle, for example a graviton scatters forward. Some of the diagrams for this are illustrated in Fig. 5. State f the same state as g; so another graviton comes in and is scattered forward. In other words we do the forward scattering of an extra graviton. In addition, from breaking matter lines we have terms for the forward scattering of an extra positron, plus the forward scattering of an extra electron, and so on; one adds the forward scattering of every possible extra particle together. That is the first contribution. But when you break up the trees, you also sometimes break two lines, and then you get diagrams like Fig. 6 with two extra particles scattering (here a graviton and electron) so it turns out you must now subtract all the diagrams with two extra particles of all kinds scattering. Then add all diagrams with 3 extra particles scattering and so on. It's a nice rule, its's quite beautiful; it took me quite a while to find; I have other proofs for orther cases that are easy to understand.

Now, the next thing that anybody would ask which is a natural, interesting thing to ask, is this. Is it possible to go back and to find the rule by which you could have integrated the closed rings directly? In other words, change the rule for integrating the closed rings, so that when you integrate them in a more natural fashion, with the new method, it will

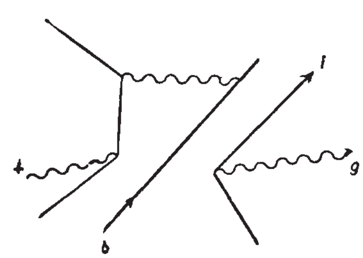


Fig. 6

give the same answer as this unique, absolute, definite thing of the trees. It's not necessary to do this, because, of course, I've defined everything; but it's of great interest to do this, because maybe I'll understand what I did wrong before. So I investigated that in detail. It turns out there are two changes that have to be made — it's a little hard to explain in terms of the gravitation of which I'll only tell about one. Well, I'll try to explain the other, but it might cause some confusion. Because I have to explain in general what I'm doing when I do a ring. Most what it corresponds to is this: first you subtract from the Lagrangian this

$$\int \sqrt{g}\, \bar{H}^{\mu\nu}{}_{;\nu} H^{\sigma}_{\mu;\sigma}\, d\tau.$$

In that way the equation of motion that results is non-singular any more. Let me write what it really is so that there's no trouble. You say to me what is this, there's a g in it and an H in it? Yes. In doing a ring, there's a field variation over which you're integrating, which I call H; and there's a g — which is the representative of all the outside disturbances which can be summarized as being an effective external field g. And so you add to the complicated Lagrangian that you get in the ordinary way an extra term, which makes it no longer singular. That's the first thing; I found it out by trial and error before, when I made it gauge invariant. But then secondly, you must subtract from the answer, the result that you get by imagining that in the ring which involves only a graviton going around, instead you calculate with a different particle going around, an artificial, dopey particle is coupled to it. It's a vector particle, artificially coupled to the external field, so designed as to correct the error in this one. The forms are evidently invariant, as far as your g-space is concerned; these are like tensors in the g world; and therefore it's clear that my answers are gauge invariant or coordinate transformable, and all that's necessary. But are also quantum-mechanically satisfactory in the sense that they are unitary.

Now, the next question is, what happens when there are two or more loops? Since I only got this completely straightened out a week before I came here, I haven't had time to inwestigate the case of 2 or more loops to my own satisfaction. The preliminary investigations that I have made do not indicate that it's going to be possible so easily gather the things into the right barrels. It's surprising, I can't understand it; when you gather the trees into processes, there seems to be some loose trees, extra trees. I don't understand them at the moment, and I therefore do not claim that this method of quantization can be obviously and evidently carried on to the next order. In short, therefore, we are still not sure, of the radiative corrections to the radiative corrections to the Lamb shift, the uncertainty lies in energies of the order of magnitude of 10^{-255} rydbergs. I can therefore relax from the problem, and say: for all practical purposes everything is all right. In the meantime, unfortunately, although I could retire from the field and leave you experts who are used to working in gravitation to worry about this matter, I can't retire on the claim that the number is so small and that the thing is now really irrational, if it was not irrational before. Because, unfortunately, I also discovered in the process that the trouble is present in the Yang-Mills theory; and secondly I have incidentally discovered a tree-ring connection which is of very great interest and importance in the meson theories and so on. And so I'm stuck to have to continue this investigation, and of course you all appreciate that this is the secret reason for doing any work, no matter how absurd and irrational and academic it looks; we all realize that no matter how small a thing is, if it has physical interest and is thought about carefully enough, you're bound to think of something that's good for something else.

DISCUSSION

Møller: May I, as a non-expert, ask you a very simple and perhaps foolish question. Is this theory really Einstein's theory of gravitation in the sense that if you would have here many gravitons the equations would go over into the usual field equations of Einstein?

Feynman: Absolutely.

Møller: You are quite sure about it?

Feynman: Yes, in fact when I work out the fields and I don't say in what order I'm working, I have to do it in an abstract manner which includes any number of gravitons; and then the formulas are definitely related to the general theory's formulas; and the invariance is the same; things like this that you see labelled as loops are very typical quantum-mechanical things; but even here you see a tendency to write things with the right derivatives, gauge invariant and everything. No, there's no question that the thing is the Einsteinian theory. The classical limit of this theory that I'm working on now is a non-linear theory exactly the same as the Einsteinian equations. One thing is to prove it by equations; the other is to check it by calculations. I have mathematically proven to myself so many things that aren't true. I'm lousy at proving things — I always make a mistake. I don't notice when I'm doing a path integral over an infinite number of variables that the Lagrangian does not depend upon one of them, the integral is infinite and I've got a ratio of two infinities and I could get a different answer. And I don't notice in the morass of things that something, a little limit or sign, goes wrong. So I always have to check with calculations; and I'm very poor at calculations — I always get the wrong answer. So it's a lot of work in these things. But I've done two things. I checked it by the mathmatics, that the forms of the mathematical equations are the same; and then I checked it by doing a considerable number of problems in quantum mechanics, such as the rate of radiation from a double star held together by quantum-mechanical force, in several orders and so on, and it gives the same answer in the limit as the corresponding classical problem. Or the gravitational radiation when two stars — excuse me, two particles — go by each other, to any order you want (not for stars, then they have to be particles of specified properties; because obviously the rate of radiation of the gravity depends on the give of the starstides are produced). If you do a real problem with real physical things in in then I'm sure we have the right method that belongs to the gravity theory. There's no question about that. It can't take care of the cosmological problem, in which you have matter out to infinity, or that the space is curved at infinity. It could be done I'm sure, but I haven't investigated it. I used as a background a flat one way out at infinity.

Møller: But you say you are not sure it is renormalizable.

Feynman: I'm not sure, no.

Møller: In the limit of large number of gravitons this would not matter?

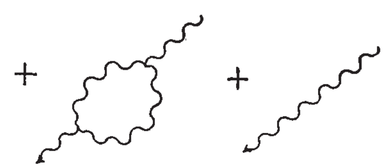
Feynman: Well, no; you see, there is still a classical electrodynamics; and it's not got to do with the renormalizability of quantum electrodynamics. The infinities come in different places. It's not a related problem.

Rosen: I'm not sure of this, not being one of the experts; but I have the impression that because of the non-linearity of the Einstein equations there exists a difficulty of the following kind. If the linear equations have a solution in the form of an infinite plane mono-chromatic wave, there does not seem to correspond to that a more exact solution; because you get piling up of energies in space and the solution then diverges at infinity. Could that have any bearing on the accuracy of this kind of calculation?

Feynman: No, I take that into account by a series of corrections. A single graviton is not the same thing as an infinite gravitational wave, because there's a limited energy in it. There's only one $\hbar\omega$.

Rosen: But you're using a momentum expansion which involves infinite waves.

Feynman: Yes, there are corrections. You see what happens if one calculates the corrections. If you have here a graviton coming in this way, then there are corrections for such a ring as this and so on. And these produce first, a divergence as usual; but second, a term in the logarithm of q^2; which means that if this thing is absolutely a free plane



wave, there's no meaning to the correction. So it must be understood in this way, that the thing was emitted some time far in the past, and is going to be absorbed some time in the future; and has not absolutely been going on forever. Then there's a very small coefficient in front of the logarithm and then for any reasonable q^2, like the diameter of the universe or something, I can still get a sensible answer; this is the shadow of the phenomenon you're talking about, that the corrections to the propagation of a graviton, dependent on the logarithm of the momentum squared carried by the graviton and which would be infinite if it were really a zero momentum graviton exactly. And so a free graviton just like that does not quite exist. And this is the correction for that. Strictly we would have to work with wave packets, but they can be of very large extent compared to the wave length of the gravitons.

Anderson: I'd like to ask if you get the same difficulty in the electromagnetic case that you did in the Yang-Mills and gravitational cases?

Feynman: No, sir, you do not. Gauge invariance of diagrams such as Fig. 2 (there is no $2D$) is satisfied whether b is a free wave or not. That is because photons are not the source of photons; they are uncharged.

Anderson: The other thing I would like to suggest is that in putting of things into baskets, you might be able to get easily by always only starting out with vacuum diagrams and opening those successively.

Feynman: I tried that and it didn't go successfully.

Ivanenko: If I understood you correctly, you had used in the initial presentation the transmutation of two particles into gravitons. Yes?

Feynman: It was one of the examples.

Ivanenko: Yes. This process was considered, perhaps in a preliminary manner, by ourselves and by Prof. Weber and Brill. I ask you two questions. Do you possess the effective cross-section? Can you indicate the effects for which high-energy processes play an important role?

Feynman: I never went to energies more than one billion-billion BeV. And then the cross-sections of any of these processes are infinitesimal.

Ivanenko: They increase very, very sharply with energy. Yes, because the radiation is quadrupole, so it increases sharply in contrast to the electromagnetic transmutation of an electron-positron pair.

Feynman: It increases very sharply indeed. On the other hand, it starts out so low that one has to go pretty far to get anywhere. And the distance that you have to go is involved in this thing — the thing that's the analogue of $e^2/\hbar c$ in electricity, which is 1/137 is non-existent in gravitation; it depends on the problem; this is so because of the dimensions of G. So if E is the energy of some process, then if you take $GE^2/\hbar c$ you get an equivalent to this $e^2/\hbar c$. It may be less than that, but at least it can't be any bigger than this. So in order to make this thing to be of the order of 1%, in which case the rate is similar to the rate of photon annihilation, at ordinary energies, we need the GE^2 to be of the order of hc, and as has been pointed out many times, that's an energy of the order 10^{-5} grams, which is 10^{18} BeV. You can figure out the answer right away; just take the energy that you are interested in, square, multiply by G and divide by hc; if that becomes something, then you're getting somewhere. You still might not get somewhere, because the cross-section might not go up that fast, but at least it can't get up any worse than that. So I think that in order to get an appreciable effect, you've got to go to ridiculous energies. So you either have a ridiculously small effect or a ridiculous energy.

Weber: I have a cross-section which may be a partial answer to Ivanenko's question. Could I write it on the board? We have carried out a canonical quantization, which is not as fancy as the one you have just heard about; but considering the interaction of photons and gravitons; and it turns out that even in the linear approximation that one has the possibility of the graviton production by scattering of photons in a Coulomb field. And the scattering cross-section for this case turns out to be $8\pi^2$ times the constant of gravitation times the energy of the scatterer times the thickness of the scatterer in the direction of propagation of the photon through it divided by c^4. This assumes that all of the dimensions of the scatterer are large in comparison with the wave length of the photon. We obtained this result by quantization, and noticed that it didn't have Planck's constant in it, so we turned around and calculated it classically. Now, if one puts numbers in this, one finds that the scattering cross-section of a galaxy due to a uniform magnetic field through it is 10^{28} cm^2, a much larger number than the object that you talked about. This represents a conversion of photons into gravitons of about 1 part in 10^{16}. This is of course too small to measure. Also, we considered the possibility of using this cross-section for a laboratory experiment in which one had a scatterer consisting, say of a million gauss magnetic field over something like a cubic meter. This turned out to be entirely impossible, a result in total contradiction to what has appeared in the Russian literature. In fact, the theory of fluctuations shows that for a laboratory experiment involving the production of gravitons by scattering of photons in a Coulomb field, the scattered power has to be greater than twice the square root of kT times the photon power divided by the averaging time of the experiment. I believe that the incorrect results that have appeared in the literature have been due to the statement that ΔP has to be greater than kT over τ; dimensionally these things are the same, but order of magnitude-wise this kind of experiment for the scatterer of which I spoke requires something like 10^{50} watts. Maybe I can say something about this afternoon; I don't want to take any more time.

De Witt: I should like to ask Prof. Feynman the following questions. First, to give us a careful statement of the tree theorem; and then outline, if he can to a brief extent, the nature of the proof of the theorem for the one-loop case, whish I understand does work. And then, to also show in a little bit more detail the structure and nature of the fictitious particle needed if you want to renormalize everything directly with the loops. And if you like, do it for the Yang-Mills, if things are prettier that way.

Feynman: I usually don't find that to go into the mathematical details of proofs in a large company is a very effective way to do anything; so, although that's the question that you asked me — I'd be glad to do it — I could instead of that give a more physical explanation of why there is such a theorem; how I thought of the theorem in the first place, and things of this nature; although I do have a proof — I'm not trying to cover up.

De Witt: May we have a statement of the theorem first?

Feynman: That I do not have. I only have it for one loop, and for one loop the careful statement of the theorem is... — look, let me do it my way. First — let me tell you how I thought of this crazy thing. I was invited to Brussels to give a talk on electrodynamics — the 50th anniversary of the 1911 Solvay Conference on radiation. And I said I'd make believe I'm coming back, and I'm telling an imaginary audience of Einstein, Lorentz and so on what the answer was. In other words, there are going to be intelligent guys, and I'll tell them the answer. So I tried to explain quantum electrodynamics in a very elementary way, and started out to explain the self-energy, like the hydrogen Lamb shift. How can you explain the hydrogen Lamb shift easily? It turns out you can't at all — they didn't even know there was an atomic nucleus. But, never mind. I thought of the following. I would explain to Lorentz that his idea that he mentioned in the conference, that classically the electromagnetic field could be represented by a lot of oscillators was correct. And that Planck's idea that the oscillators are quantized was correct, and that Lorentz's suggestion, which is also in that thing, that Planck should quantize the oscillators that the field is equivalent to, was right. And it was really amusing to discover that all that was in 1911. And that the paper in which Planck concludes that the energy of each oscillator was not $n\hbar\omega$ but $(n+1/2)\hbar\omega$ which was also in that, was also right; and that this produced a difficulty, because each of the harmonic oscillators of Lorentz in each of the modes had a frequency of $\hbar\omega/2$ which is an infinite amount of energy, because there are an infinite number of modes. And that that's a serious problem in quantum electrodynamics and the first one we have to remove. And the method we use to remove it is to simply redefine the energy so that we start from a different zero, because, of course, absolute energy doesn't mean anything. (In this gravitational context, absolute energy does mean something, but it's one of the technical points I can't discuss, which did require a certain skill to get rid of, in making a gravity theory; but never mind.) Now look — I make a little hole in the box and I let in a little bit of hydrogen gas from a reservoir; such a small amount of hydrogen gas, that the density is low enough that the index of refraction in space differs from one by an amount proportional to A, the number of atoms. With the index being somewhat changed, the frequency of all the normal modes is altered. Each normal mode has the same wavelenght as before, because it must fit into the box; but the frequencies are all altered. And therefore the $\hbar\omega$'s should all be shifted a trifle, because of the shift of index, and therefore there's a slight shift of the energy. Although we subtract $\hbar\omega/2$ for the vacuum, there's a correction when we put the gas in; and this correction is proportional to the number of atoms, and can be associated with an energy for each atom. If you say, yes, but you had that energy already when you had the gas in back in the reservoir, I say, but let us only compare the difference in energy between the $2S$ and $2P$ state. When we change the excitation of the hydrogen gas from $2S$ to $2P$ then it changes its index without removing anything; and the energy difference that is needed to change the energy from $2S$ to the $2P$ for all these atoms is not only the energy that you calculate with disregard of the zero point energy; but the fact is that the zero point energy is changed very slightly. And this very slight difference should be the Lamb effect. So I thought, it's a nice argument; the only question is, is it true. In the first place it's interesting, because as you well know the index differs from one by an amount which is proportional to the forward scattering for γ rays of momentum k and therefore that shift in energy is essentially the sum over all momentum states of the forward scattering for γ rays of momentum k. So I looked at the forward scattering and compared it with the right formula for the Lamb shift, and it was not true, of course; it's too simple an argument. But then I said, wait, I forgot something. Dirac, explained to us that there are negative energy states for the electron but that the whole sea of negative energy states is filled. And, of course, if I put the hydrogen atoms in here all those electrons in negative energy states are also ascattering off the hydrogen atoms; and therefore their states are all shifted; and therefore the energy levels of all those are shifted a tiny bit. And therefore there's shift in the energy due to those. And so there must be an additional term which is the forward scattering of positrons, which is the same as scattering of negative energy electrons. Actually, for the symmetry of things it is better to take half the case where you make the positrons the holes and the other half where you make the electrons the holes; so it should be 1/2 forward scattering by electrons, 1/2 scattering by positrons and scattering by γ rays — the sum of all those forward scattering amplitudes ought to equal the self--energy of the hydrogen atom. And thats' right. And it's simple, and it's very peculiar. The reason it's peculiar is that these forward scatterings are real processes. At last I had discovered a formula I had always wanted, which is a formula for energy differences (which are defined in terms of virtual fields) in terms of actual measurable quantities, no matter how difficult the experiment may be — I mean I have to be able to scatter these things. Many times in studying the energy difference due to electricity (I suppose) between the proton and the neutron, I had hoped for a theorem which would go something like this — this energy difference between proton and neutron must be equal to the following sum of a bunch of cross-sections for a number of processes, but all real physical processes, I don't care how hard they are to measure. So this is the beginning of such a formula. It's rather surprising.

It's not the same as the usual formula — it's equal to it but it's not the same. I have no formulation of the laws of quantum gravidynamics; I have a proposal on how to make the calculations. When I make the proposal on how to do the closed loops, the obvious proposal does not work; it gives non-unitarity and stuff like that. So the obvious proposal is no good; it works O.K. for trees; so how am I going to define the answer for would correspond to a ring? The one I happen to have chosen is the following: I take the ring in general for any meson theory, one closed ring can be written as equivalent to a whole lot of processes each one of which is trees. I then define, as my belief as to what the ring ought to be in the grand theory, that it's going to be also equal to the corresponding physical set of trees. When I said this is equal to this. I didn't worry about gauge or anything else; what I means was, if these weren't gravitons but photons or any other neutral object — it doesn't make any difference what they are — this theorem is right. So I suppose it's right also for real gravitons, and I suppose also that what's being scattered is only transverse and is only a real free graviton with $q^2 = 0$. Therefore, I say let this ring equal this set of trees. Every one of these terms can be completely computed — it's a tree. And it's gauge invariant; that is, if I added an extra potential on the whole thing, another outside disturbance of a type which is nothing but a coordinate transformation — in short a pure gradient wave — to the whole diagram then it comes on to all of these processes; but it makes no effect on any of them, and therefore makes no effect on the sum; and therefore I know my definition of this ring is gauge-invariant. Second, unitarity is a property of the breaking of this diagram; the imaginary part of this equals something; if you take the imaginary part of this side, it's already broken up, in fact, and you can prove immediately that it's the correct unitarity rule. Therefore it's going to be unitarity and so on and so on. And so I therefore define gravity with one ring in this way. Now what prevents me from doing it with two rings? The lack of a complet statement of what two rings is equal to in terms of processes; that is I can open the ring all right; but I can't put the pieces — the broken diagrams — back together again into complete sets that each one is a complete physical process. In other words some of them correspond to the scattering of a graviton, but leaving out some diagrams. But the scattering of a graviton leaving out diagrams is no longer gauge invariant, I mean, not evidently gauge invariant, and so the power of the whole thing collapses. I don't know what to do with it. So that's the situation; that's why it is crucial to the particular plan. There's always, of course, another way out. And that's the following (and that's what I tried to describe at the end of the talk — maybe I talked too fast): After all now I've defined what this results is equal to — by definition not that you should do a loop some way and get this, but that a loop is equal to this by defition, and I'm not going to do a loop any other way. But, of course, from a practical point of view or from the point of view purely of interest, the question is, can you come back now and calculate the ring directly by some particular mathematical shenanigans, and get the same answer as you get by adding the trees. And I found the way to do that. I have another way, in other words, to do the ring integral directly. I have to subtract something from a vector particle going around the instead of a graviton to get the answer right. So I known the rule, and I know why the rule is, and I have a proof of the rule for one loop. I have two ways of extending. I can either break this two loop diagram open and get it back into the processes, like I did with the one ring — where so far I'm stuck. Or, I can take the rule which I found here and try to guess the generalization for any number of rings. Also stuck. But I've only had a week, gentlemen; I've only been able to straighten out the difficulty of a single ring a week ago when I got everything cleaned up. It's more than a week — I had to take a lot of time checking and checking; but I was only finished checking to make sure of everything for this conference. And of course you're always asking me about the thing I haven't had time to make sure about yet, and I'm sorry; I worked hard to be sure of something, and now you ask me about those things I haven't had time. I hoped that I would be able to get it. I still have a few irons to try; I'm not completely stuck—maybe.

DeWitt: Because of the interest of the tricky extra particle that you mentioned at the end, and its possible connection, perhaps, with some work of Dr Białynicki-Birula, have you got far enough on that so that you could repeat it with just a little more detail? The structure of it and what sort of an equation it satisfies, and what is its propagator? These are technical points, but they have an interest.

Feynman: Give me ten minutes. And let me show how the analysis of these tree diagrams, loop diagrams and all this other stuff is done mathematical way. Now I will show you that I too can write equations that nobody can understand. Before I do that I should like to say that there are a few properties that this result has that are interesting. First of all in the Yang-Mills case there also exists a theory which violates the original idea of symmetry of the isotopic spin (from which was originally invented) by the simple assumption that the particle has a mass. That means to add to the Lagrangian a term $-\mu^2 a_\mu a^\mu$ where a_μ is an isotopic vector. You add this to the Lagrangian. This destroys the gauge invariance of the theory — it's just like electrodynamics with a mass, it's no longer gauge-invariant, it's just a dirty theory. Knowing that there is no such field with zero mass people say: „let's put the mass term on". Now when you put a mass term on it is no longer gauge invariant. But then it is also no longer singular. The Lagrangian is no longer singular for the same reason that it is not invariant. And therefore everything can be solved precisely. The propagator instead of being $\delta_{\mu\nu}$ between two currents is

$$\frac{\delta_{\mu\nu}-q_\mu q_\nu/\mu^2}{q^2-\mu^2}, \tag{10}$$

where q_μ is the momentum of propagating particle. The factor $1/(q^2-\mu^2)$ is typical for mass μ but the part $-q_\mu q_\nu/\mu^2$ is an important term which can be taken to be zero in electrodynamics but it is not obvious whether it can be taken to be zero in the case of Yang-Mills theory. In fact it has been proved it cannot be taken to be zero; this propagator is used between two currents. I am using the Yang-Mills example instead of the gravity example. I really want only the case $\mu^2 = 0$, and am asking whether I can get there by first calculating finite μ^2, then taking the limit $\mu^2 = 0$.

Now, with $\mu^2 \neq 0$ this is a definite propagator and there are no ambiguities at the closed rings, the closed loops. I have no freedom, I must compute this propagator. I mean there is no reason for trouble, and there is no trouble. There is no gauge invariance either.

And of course I checked. I broke the rings and I computed by the broken ring theorem method a closed loop problem of fair complexity (which in fact was the interaction of two electrons). I computed it by the open ring method and by the closed ring method, and of course it agreed, there is no reason that it shouldn't. It turned out that for tree diagrams you don't have to worry about this $q_\mu q_\nu/\mu^2$ term, you can drop it — but not for the closed ring — only for tree. Therefore the tree diagrams have the definite limit as μ^2 goes to zero. And yet I have the closed ring diagram which is equal to the tree diagram when the mass is anything but zero, and therefore it ought to be true that the limit as μ^2 goes to zero of the ring is equal to the case when $\mu = 0$. It sounds like a great idea why don't you define the desired $\mu^2 = 0$ theory that way? Answer: You can't put μ^2 equal zero in the form (10). You can't do it because of the $q_\mu q_\nu/\mu^2$. So it was necessary next to see if there is a way to re-express the ring diagrams, for the case with $\mu^2 \neq 0$, in a new form with a propagator different from (10), that didn't have a μ^2 in it, in such a form that you can take the limits as μ^2 goes to zero. Then that would be a new way to do the μ equal zero case; and that's the way I found the formula. I'll try to explain how to find that theory.

We start with a definite theory, the Yang-Mills theory with a mass (the reason I do that is that there's no ambiguity about what I am trying to do) and later on I take the mass to zero, then the theory works something like this. You have the Lagrangian $\mathcal{L}(A, \varphi)$ which involves the vector potential of this field and the fields φ representing the matter with which this object is interacting for zero mass, to which, for finite mass we add the term $\mu^2 A_\mu A_\mu$. This is the Lagrangian that has to be integrated and the idea is that you integrate this over all fields A and φ; and that is the answer for the amplitude of the problem

$$X = \int e^{\int \mathcal{L}(A,\varphi)d\tau + \mu^2 A_\mu A_\mu d\tau}\, DAD\varphi. \tag{11}$$

But wait, what about the initial and final conditions? You have certain particles coming in and going out. To simplify things (this is not essential) I'll just study the case that corresponds only to gravitons in and out. I'll call them gravitons and mesons even though they are vector particles. The question is first, what is the right answer if you have gravitons represented by plane waves, A_1, A_2, A_3 ... going in (positive frequency in A_1) or out (negative frequency). You make the following field up. Let A_{asym} be defined as α times the wave function A_1 that represents the first graviton coming in a plane wave, plus β times A_2 plus γ times A_3 and so on.

$$A_{\text{asym}} = \alpha A_1 + \beta A_2 + \gamma^{A_3} \ldots,\ A \to A_{\text{asym}}. \tag{12}$$

Then you calculate this integral (11) subject to the condition that A approaches A_{asym} at infinity. The result of this is of course a function of α, β, γ ... and so on. Then what you want for X is just the term first order in α, β, γ ... That means just one of each these gravitons coming in and out. That's the right formula for a regular theory, for meson theory, You calculate the integral subject to the asymptotic condition, when you imagine all these waves, but you take the first order perturbation with respect to each one of the incoming waves. You never let the same photon operate twice; a photon operating twice is not a photon, it is a classical wave. So you take the derivative of this with respect to α, β, γ and so on, then setting them all equal to zero. That's problem. (In general there's φ asymptotic too.)

Now the way I happened to do this is the following: Let us call A_0 the A which satisfies the classical eqatiuons of motion, which in this particular case will be

$$\left.\frac{\partial \mathcal{L}}{\partial A}\right|_{A^\circ} + \mu^2 A^\circ = 0 \tag{13}$$

I solve this subject to the condition that A_0 equals A_{asym}. In other words, I find what is the maximum or minimum — whatever it is — of the action in (11), subject to the asymptotic condition. That's the beginning of analysing this.

The next thing is to make the simple substitution $A = A_0 + B$ and put it back in equation (11). Then if you take $\mathcal{L}$ od A_0+B (if B is negligible you get $\mathcal{L}$ of A_0 and so forth) so you get something like this

$$e^{i[\mathcal{L}(A_0)+\mu^2 A_0 A_0]} \int e^{[\mathcal{L}(A^0+B)-\mathcal{L}(A^0)]+\mu^2 BB+2\mu^2 AB}\, DB. \tag{14}$$

The integral is over all B, and B must go to zero asymptotically. This business can be expanded in powers of B.

$$\mathcal{L}(A+B)-\mathcal{L}(A)+\mu^2 BB+2\mu^2 AB = \text{Quad}\,(B)+\text{Cubic}\,(B)+\ldots+\mu^2 BB. \tag{15}$$

The zeroth power B is evidently zero. The first power of B is also zero because A_0 minimized the original thing. So this starts out quadratic in B plus cubic in B plus *etc.*, that's what this is here. These quadratic forms Quad (B) and so on of course depend on A_0, the cubic form involves A_0 in some complex, maybe very complicated, locked-up mess, but as far as B is concerned it is second power and higher powers.

Now I would like to point something out. First — it turns out if you analyze it, that the contribution of the first factor here alone (if you had forgotten the intergal and called it one) is exactly the contribution of all trees to the problem. So that's like the classical theories related to trees. Next, if you drop the term cubic in B in the exponent completely and just integrated the result over DB, that corresponds to the contribution from one ring, or from two isolated rings, or three isolated rings, but not interlocked rings. If you start to include the cubic term is has to come in a second power to do anything, because of the evenness and oddness of function. And as soon as it comes in second power, the cubic term, having three of these things come together twice, makes a terrible thing like ∞ which is a double ring. So you don't get to a double ring until you bring a cubic term down to the second order. So if I disregard that and just work with this second order term Quad (B) $+\mu^2 BB$, I'm studying the contribution from one ring. If I study this I am working from the trees. And now you see I have in my hands an expression for the contribution of a ring correct in all orders no matter how many lines come in. I also have expressions for the contributions from trees and so on. I can compare them in different mathematical circumstances, and it's on this basis that I have been able to prove everything I have been able to prove relating one ring to trees.

Now, let me explain how the theorem was obtained that takes the case for the mass and for a ring. Now we have to discuss a ring, which is a formula like this

$$X = \int e^{(\text{Quadr}(B)+\mu^2 B^2)}\, DB. \tag{15}$$

The quadratic form involves A_0 so the answer depends on A_0 — it's some complicated functional of A_0. Anyway I won't say that all the time, I'll just remember that. We have to integrate over all B. And the difficulty is — not difficulty, but the point is — that this quadratic form in B is singular, because it came from the piece of the action that has an invariance and this invariance keeps chasing us along. And there are certain transformations of B which leave this Quard B part unchanged in first order. That transformation in the Yang-Mills theory is

$$\vec{B}'_\mu = \vec{B}_\mu + \vec{\nabla}\alpha + (\vec{\alpha}\times\vec{A}) = \vec{B}_\mu + \alpha_{;\mu}\,. \tag{16}$$

where the vectors are in isotopic spin space and α is considered as first order. This transformation leaves the quadratic form invariant so the Quad (B) thing by itself is singular. But it doesn't make any difference, because of the addition of the μ^2BB. If $\mu^2 \neq 0$, there is no problem, but if $\mu^2 \to 0$, I'd be in trouble.

I discovered that if I make this change (16) in the actual Lagrangian and carry everything up to second order it is exact, in fact because it's only second order. If I do it with the exact change, the thing isn't invariant, it is only invariant to first order in α. But if I make the substitution exactly, then I get a certain addition to the Lagrangian, in other words the Lagrangian of B' (this includes the μ^2, the Lagrangian plus the μ^2 term in B) is the Lagrangian plus the μ^2 term in B plus something like this

$$\mu^2 B_\mu \cdot \alpha_{,\mu} + \frac{1}{2}\,\mu^2 \alpha_{,\mu}\,\alpha_{;\mu}$$

I have to explain that the semicolon is analogous to the semicolon in gravity. The semicolon derivative $X_{;\mu}$ means the ordinary derivative of X minus A cross X and that's the analogue of the Christoffel symbols. Anyway, I find out what happens to L when I make this transformation. Now comes the idea, the trick, the nonsense: you start with the following thing; you, say, suppose instead of writing the original terms down, instead of writing the original Lagrangian I were to write the following:

$$\int e^{\mathcal{L}(B) + \frac{1}{2}(B_{\mu,\mu} - \alpha_{;\mu\mu} + \mu^2\alpha)^2}\,\mathcal{D}\alpha\mathcal{D}B.$$

Now I say that the integral over α is some constant or other. So all I have done is to multiply my original integral by $\mathcal{L}$ of B (by $\mathcal{L}$ of B I mean the whole thing, I mean this whole thing is going to be $\mathcal{L}$ of B). If I can claim that when I integrate α I get something which is independent of B, which is not self-evident. If I integrate over all α it does not look as if it is independent of B — but after a moment's consideration you see that it is. Because if I can solve a certain equation, which is $\alpha^{\mu}_{;\mu} - \mu^2\alpha = B^{\mu}_{,\mu}$, I can shift the value of α by that amount, and then this term would disappear. In other words if I can solve this, and call this solution α_0 and change α to α_0, then the B would cancel and it would only be α' here. I did it a little abstractly which is a little easier to explain, therefore, this term that I've added can be thought of as an integral of the following nature: Integral of some B, plus an operator acting on α (this complicated operator is the second derivative and so on) squared $\mathcal{D}\alpha$. And then by that substitution I've just mentioned, this becomes equal to 1/2 the operator on A

times α' squared $\mathcal{D}\alpha$, which is equal to the integral e to the one half of α times A, the operator A, times the operator A times α integrated over primed α. Now when you integrate a quadratic form, which is a quadratic with an operator like this you get one over the square root of the determinant of the operator. So this thing is one over the square root of the determinant of the operator AA. The determinant of the operator A times A is square of the determinant of A. So this is one over the determinant of the operator A, or better it is one over square root of the determinant of the operator A squared, you'll see in a minute why I like to write it in this way. In other words, when I've written this thing down I've written the answer that I want. Let's call X the unknown answer that I want. Then this is equal to X divided by this determinant's square root squared. Now comes the trick — I now make the change from B to B'. We notice that B changed to B' is simply... oh!, this is wrong, that's what's wrong, it should be just this. Now I've got it. The change from B' to B is to add something to B. Therefore to the differential of B it adds nothing, it's just shifting the B to a new value. So I make the transformation from B to B' everywhere. So then I have $d\alpha$ and dB, and now I have a new thing up here where I make use of the formula for $\mathcal{L}$ of B':

$$\mathcal{L}(B') = \mathcal{L}(B) + \mu^2 B_\mu \alpha_\mu + \frac{1}{2}\,\mu^2 \alpha_{,\mu}\alpha_{;\mu}$$

You see there is a certain cross term generated here and another cross term coming from expanding this out and the net result, with a little algebra here, is that becomes $\mathcal{L}$ of B, but the quadratic term doesn't cancel out and is left; there's one half of $B_{\mu,\mu}$ squared; that's from this term; the cross term here cancels the cross term in there; and then we have only the quadratic — I mean the α terms

$$\int e^{\mathcal{L}(B)+\frac{1}{2}(B_{\mu,\mu})^2}\,\mathcal{D}B\, e^{\frac{\mu^2}{2}(\alpha_{,\mu}\alpha_{;\mu}+\mu^2\alpha^2)}\,\mathcal{D}\alpha.$$

And the problem is now to do this integral on α; well, another miraculous thing happens. I have the operator A, but that this down thing is $\alpha A\alpha$, and therefore its result is just determinant once; or the square of this integral is equal to this determinant, or something like that. Therefore, when you get all the factors right, X, the unknown, is equal to

$$X = \Big[\int e^{\mathcal{L}(B)+\frac{1}{2}(B_{\mu,\mu})^2}\,\mathcal{D}B\Big] : \Big[\int e^{\frac{\mu^3}{2}(\alpha_{,\mu}\alpha_{;\mu}+\mu^2\alpha^2)}\,\mathcal{D}\alpha\Big].$$

Sachs: I want to ask a question about long-range hopes. Perhaps for irrational reasons people are particularly interested in those parts of the theory where is a possibility of real qualitative differences: what do the coordinates or topology mean in a quantized theory, and this kind of junk. Now I wonder if you think that this perturbation theory can eventually be jazzed up to cover also this kind of questions?

Feynman: The present theory is not a theory as it is incomplete. I do not give a rule on how to do all problems. I expect of course that if I spend more time on figuring out how to untangle the pretzels I shall be able to make it into such a theory. So let's suppose I did. Now you can ask the question would the completed job, assuming it exists, be of any interest to esoteric question about the quantization of gravity. Of course it would be, because it would be the expression of the quantum theory; there is today no expression of the quantum theory which is consistent. You say: but it's perturbation theory. But it isn't. I worked on the thing analyzing it in the series of increasing accuracy, but that's only, obviously, when I am doing problems and checking, or doing things like I just did. But even there I haven't said how many times the vector potential A_0 is attacking the diagram, there is no limit to what order of external lines are involved in the calculation of A_0, for example. And so if I get my general theorem for all orders, I'll have some kind of a formulation. The fact is, that in such things as electrodynamics and other theories, it has not been possible to figure out the consequences of the quantum field theory in the case of strong interactions, because of technical difficulties which are not technical difficulties just of the gravitation theory, but exist all over the quantum field theory. I do not expect that the gravitational problems will be any easier in that region than they are in any other field theory, so I can say very little there. But at least one should certainly formulate the theory that you're trying to calculate first, and then find out what the consequences are, before trying to do it the other way round. So I think that you'll be frustrated by the difficulties that do appear whenever any theory diverges. On other hand, if you ask about the physical significance of the quantization of geometry, in other words about the philosophy behind it; what happens to the metric, and all such questions, those I believe will be answerable, yes. I think you would be able to figure out the physics of it afterwards, but I won't to think about that until I have it completely formulated, I don't want to start to work out the anser to something unless I know what the equation is I am trying to analyze. But I dont' have the doubt that you will be able to do something, because after all you are describing the phenomena that you would expect, and if you deseribe the phenomena then you expect you can then find some kind of framework in which to talk to help to understand the phenomena.

remarks on twistors and curved-space quantization. *Quantum gravity 2: A second* *d Symposium* Eds C. J. Isham, R. Penrose, and D. W. Sciama, 578–92. 1981

ıd, G.B. and Kohn, J.J. (1972). *The Neumann Problem for the Cauchy-ıann Complex*. Princeton University Press.
....ıı, C.D. and MacKichan, B. (1977). *Ann. Scu. norm. sup. Pisa* CI. *Sci.* **4**, 577.
Hughston, L.P. and Ward, R.S. (Eds) (1979). *Advances in Twistor Theory*. Pitman, London.
Morrow, J. and Kodaira, K. (1971). *Complex Manifolds*. Holt, Rienhard, and Winston, New York.
Nirenberg, L. (1973). *Lectures on Partial Differential Equations*. Amer. Math. Soc., Providence, Rhode Island.
Penrose, R. (1967). *J. Math. Phys.* **8**, 345.
—— (1968). In *Battelle Rencontres*. (ed. C.M. DeWitt and J.A. Wheeler). Benjamin, New York.
—— (1969). *J. Math. Phys.* **10**, 38.
—— (1975). In *Quantum Gravity, an Oxford Synposium*. (ed. C.J. Isham, R. Penrose, and D.W. Sciama). Oxford University Press.
—— (1976). *Gen. rel. Grav.* **7**, 31.
—— (1979*a*). In *Complex Manifold Techniques in Theoretical Physics*. (ed. D.E. Lerner and P.D. Sommers). Pitman, London.
—— (1979*b*). In *Advances in Twistor Theory*. (ed. L.P. Hughston and R.S. Ward). Pitman, London.
—— and MacCallum, M.A.H. (1972). *Phys. Reports*. 6, 241.
—— and Ward, R.S. (1980). In *General Relativity and Gravitation, One Hundred Years after the Birth of Albert Einstein*. Vol. 2. (ed. A. Held). Plenum, New York.
Plebanski, J. (1965). *Acta Phys. Polon.* **27**, 361.
Sachs, R. (1961). *Proc. R. Soc. Lond.* **A264**, 309.
Wells, R.O., Jr. (1979). *Differential Analysis on Complex Manifolds*. Springer-Verlag, New York.
Ward, R.S. (1977). *Curved Twistor Spaces*. D. Phil. thesis, Oxford.

123
ON SCHWARZSCHILD CAUSALITY—A PROBLEM FOR "LORENTZ COVARIANT" GENERAL RELATIVITY

Essays in general relativity (A. Taub Festschrift). Ed. F. J. Tipler, 1–12. Academic Press, 1980

A colleague has told me that he believes that there is an error in one of the detailed calculations presented here. I have unfortunately not had the time to check these calculations through again since then. I have no particular reason to doubt his contention, but I cannot believe that such an error would substantively alter the basic conclusions of the article, which are, at root, based on other clear geometrical arguments. Nevertheless, I would welcome clarification on this issue.

1

On Schwarzschild Causality—A Problem for "Lorentz Covariant" General Relativity

Roger Penrose

Mathematical Institute
University of Oxford
Oxford, England
and
Department of Mathematics
University of California
Berkeley, California

I. Introduction

There appears to be a viewpoint, prevalent among some physicists [1] (cf. [2]), that while a geometrical approach to general relativity may have merits on aesthetic grounds and may have appeal for those whose interests are, perhaps, essentially pure mathematical, a strong emphasis on curved-space geometry is nevertheless to be rejected if real physical understanding and important future progress are to be achieved in gravitation theory. Abe Taub, however, is clearly not of this way of thinking. His many important contributions to both the physical and geometrical aspects of general relativity bear strong witness to the fact that far from being an obstacle to progress, differential geometry is an efficient and essentially indispensable tool in this highly significant aspect of physical insight.

As a way of honoring Abe's retirement, I shall present here a result that lends strong additional support to this geometrical viewpoint. It is directed, particularly, against the idea that general relativity might be adequately described as though it were a Lorentz-covariant (or, more correctly, Poincaré-covariant) field theory according to which the physical metric tensor is to be

treated as though it were not significantly different from any other field tensor.

II. Lorentz Covariance and Causality

Now the fundamental and unique role played by the physical metric tensor g (or at least by its conformal part $\hat{g}$, which represents 9 out of its 10 algebraically independent components) is, indeed, that it defines the *physical causal relations between points*. These causal relations play a key role in any classical relativistic field theory since they determine the propagation directions for *all* relativistic fields. Furthermore, the significance of this causal structure is as great in quantum field theory as in classical field theory. Quantum causality has the implication that field operators at spacelike-separated points must necessarily commute. If we were to take the standard "Lorentz-covariant" view, then we would need to introduce a background Minkowski metric η (perhaps not canonically) with the property that any two field operators at points that are spacelike-separated with respect to the *flat* causal structure defined by η would necessarily commute. (To modify this rule would be to reject the standard Lorentz-covariant viewpoint, from which all the standard results, such as the PCT and spin-statistics theorems are derived.) This is not to say that the final causality that is physically observed need agree with that defined by η. The actual way that fields propagate in the resulting theory would have to be calculated in detail. A normal procedure for doing this would be to obtain the metric g from a power series expansion of Lorentz-covariant terms, this being an infinite summation of Feynman graphs. (Summing "tree diagrams" is to give the classical g-field.) If such a Lorentz-covariant theory is to agree with general relativity, then the finally derived field propagation has to follow the null cones of this resulting general-relativistic curved metric g instead of those of η.

For a satisfactory theory, however, one would anticipate an important consistency requirement relating η to g: *the causality defined by g should not violate the background η-causality.* To put this another way, the g-null-cones ought never to extend outside the η-null-cones (Fig. 1). Thus, timelike curves with respect to g should always remain "timelike" with respect to η, i.e.,

$$g(T, T) > 0 \Rightarrow \eta(T, T) > 0 \quad \text{at every point,} \qquad \text{(II.1)}$$

for every tangent vector T (using the Lorentzian signature $+ - - -$). I write this condition

$$g < \eta. \qquad \text{(II.2)}$$

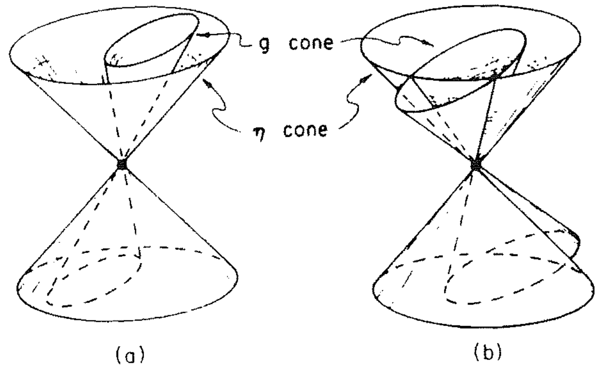


Fig. 1 Physical g-causality should not violate background η-causality. (a) Allowed: $g < \eta$ holds and (b) forbidden: $g < \eta$ fails.

If this condition were to *fail* to hold, then *physical* field propagation, which will follow the g-cones (or arbitrarily closely to them), would be superluminary with respect to η-causality. If we anticipate "physical" quantum field operators, describing the physically measurable fields, and constructible (say by infinite summation) from the original Lorentz-covariant operators, then these new field operators ought to be noncommuting for g-null- (or perhaps g-timelike-) separated points. But if $g < \eta$ fails, then these operators would have also to be noncommuting at certain pairs of η-spacelike-separated points (namely those along g-cones that extend outside η-cones). But this is not possible since noncommuting operators cannot be built out of the original commuting ones at spacelike η-separation.

Of course, in a proper quantized theory of general relativity, in which g also becomes a quantized field, the "g-cones" would never be perfectly well defined. However, if such a theory were to bear any resemblance to standard general relativity in the *classical limit*, then there should be a resulting "approximate" classical g-metric that would be anticipated to satisfy $g < \eta$, for the reasons outlined above. In the absence of a good theory of quantized gravity, there necessarily remains a certain inconclusiveness in this argument. Nevertheless, a question of some considerable interest for its *own* sake is whether or not, for a physically reasonable metric g on a space–time manifold M, a Minkowski metric η also exists on M with $g < \eta$. And if not, then this fact would also seem to provide some pertinent evidence against the fruitfulness of a Lorentz-covariant approach.

It is clear that for certain "strong gravitational fields" we should have difficulty in arranging $g < \eta$. For example, the *topology* of the space–time might itself differ from that of Minkowski space. Also there are space–times with causality violations (such as a maximally extended Kerr solution [3])

and certain examples (such as plane waves [4]), having no Cauchy hypersurfaces, and for which, for rather blatant reasons, $g < \eta$ cannot be arranged. But all these examples could be reasonably argued to be "nonphysical" and therefore not to constitute any significant case against the Lorentz-covariant viewpoint. However, I shall show here that the situation is much more serious than this. In even the simplest of physically reasonable nontrivial space–time geometries, namely, the Schwarzschild solution (with any suitable interior) $g < \eta$ indeed *cannot* be *satisfactorily* arranged.

III. Schwarzschild Geometry

At first, it would seem that the standard Schwarzschild coordinate system (t, r, θ, ϕ), for which the g-metric takes the form

$$ds^2 = \left(1 - \frac{2m}{r}\right)dt^2 - \left(1 - \frac{2m}{r}\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \quad \text{(III.1)}$$

provides a counter example to above contention, since all the null cones of this metric do in fact lie inside those of the corresponding Minkowski metric (in spherical polar coordinates):

$$d\sigma^2 = dt^2 - dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \quad \text{(III.2)}$$

provided that we restrict to the range

$$r \in [r_0, \infty) \quad \text{(III.3)}$$

for some $r_0 > 2m$. In order to cover the remaining range, we need an "interior" solution, such as the constant density perfect fluid solution originally proposed by Schwarzschild [1]

$$ds^2 = \left[\frac{3}{2}\left(1 - \frac{2m}{r_0}\right)^{1/2} - \frac{1}{2}\left(1 - \frac{2mr^2}{r_0^3}\right)^{1/2}\right]^2 dt^2 - \left[1 - \frac{2mr^2}{r_0^3}\right]^{-1} dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \quad \text{(III.4)}$$

where

$$r \in [0, r_0], \qquad r_0 > 3m. \quad \text{(III.5)}$$

(I am not concerned with the question of black holes here, which would lead to some added complications.) These two solutions (III.1) and (III.4) match together to give a manifold with a C^0-metric g for which $g < \eta$, where the flat metric is given by (II.2). [It would not be hard, in fact, to modify the interior metric (III.4) to obtain a C^∞-metric g with this property, especially if we are not concerned with any particular equation of state in the interior

CHAPTER 123

region.] We might think, therefore, that the effect of the Schwarzschild gravitational field is simply to "slow down" the velocity of light and so to provide $g < \eta$.

There is, however, an important sense in which this setup is quite unsatisfactory. This relates to behavior at infinity. One of the reasons for adopting a Lorentz-covariant viewpoint would, after all, be to discuss scattering theory. We might be concerned with incoming and outgoing asymptotically plane waves (e.g., gravitational perturbations or test Maxwell or massless scalar fields). Alternatively, taking the geometrical optics limit, we could be concerned with incoming and outgoing null geodesics. But, with regard to these null geodesics and null surfaces, the g-metric given by (III.1) actually differs very greatly from the η-metric given by (III.2) at large distances from the matter source (despite the fact that the metrics appear, naïvely, to go into one another in the limit $r \to \infty$). Consider, for example, the radial outgoing g-null geodesics given by

$$u = \text{const}, \qquad \theta = \text{const}, \qquad \phi = \text{const},$$

where u is g-retarded time:

$$u = t - r - 2m \log(r - 2m). \tag{III.6}$$

It is clear from the form of (III.6) that as $r \to \infty$, the value of $t - r$ is unbounded above along the geodesic. But $t - r$ is the standard η-retarded time, so we see that outgoing g-null geodesics reach indefinitely far into the η-retarded future, and do not correspond at all to outgoing η-null geodesics. Correspondingly, the incoming g-null geodesics are generators of the incoming g-null cones along which the g-advanced time

$$v = t + r + 2m \log(r - 2m) \tag{III.7}$$

is constant. Again there is no correspondence with the incoming η-null geodesics since the η-advanced time $t + r$ becomes unboundedly large and negative into the past along a g-null geodesic.

This is clearly unsatisfactory if this η-metric is to be used in a Lorentz-covariant approach for studying scattering theory in a Schwarzschild space time. On the other hand, we might envisage using an η-metric which is related to the Schwarzschild g-metric in a *different* way from the one just considered. To put things another way, we might choose a different coordinate system for the space-time, whose naturally associated flat metric was not the same as the η just considered. For example, we could use

$$r' = \tfrac{1}{2}(v - u) = r + 2m \log(r - 2m) \tag{III.8}$$

in place of r and now adopt the flat η'-metric given by

$$d\sigma'^2 = dt^2 - dr'^2 - r'^2(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{III.9}$$

In this particular case, since $u = t - r'$ and $v = t + r'$ are retarded and advanced times, respectively, for *both* metrics, the difficulty that was just mentioned concerning the difference in asymptotic structure between g and η does *not* now arise. However, it turns out that the condition $g < \eta$ is now what fails!

The purpose of this article is to show that there is, indeed, an essential incompatibility between the causal structure in Schwarzschild space-time (of positive mass m, and with *any* interior whatever) and that of Minkowski space. This shows up either asymptotically or in a violation of the local condition $g < \eta$. The most convenient way to handle the asymptotic conditions is by use of *conformal infinity* [3,5,6]. Both the Schwarzschild and Minkowski space-times have, in fact, well-behaved conformal infinities $\mathscr{I}^{\pm}$. Each point p of $\mathscr{I}^+$ can be thought of as describing an *outgoing (null) asymptotically plane wavefront*. (This is [3,7] the boundary $\partial I^-(p)$ of the TIP $I^-(p)$, cf. (IV. 2), representing p—an "event horizon" [6].) Likewise each point q of $\mathscr{I}^-$ describes an *incoming (null) asymptotically plane wavefront* (the boundary $\partial I^+(q)$ of the TIF $I^+(q)$ representing q—a "creation horizon" [3], sometimes referred to as a "particle horizon" [6]). Thus, $\mathscr{I}^+$ and $\mathscr{I}^-$ have very direct and natural interpretations in the context of scattering theory.

The space-time manifold with both conformal infinities $\mathscr{I}^{\pm}$ adjoined will be denoted $\overline{M}$. The conformal metric $\hat{g}$ is well-defined (C^0 will do) on the whole of $\overline{M}$, including its boundary $\mathscr{I}^- \cup \mathscr{I}^+$. The required condition that the asymptotic η-causal structure agree with the asymptotic g-causal structure (i.e., that the scattering theory be compatible for both metrics) is that both $\hat{g}$ and $\hat{\eta}$ be well defined on the *same* manifold-with-boundary $\overline{M}$. We have seen that this does *not* in fact hold for the $\hat{g}$ and $\hat{\eta}$ conformal metrics related to each other as is entailed by (III.1) and (III.2), but that it *does* hold for $\hat{g}$ and $\hat{\eta}$, related by (III.1) and (III.9).

IV. The Theorem

Let us now consider the more general question of whether or not a flat conformal $\hat{\eta}$-metric [not necessarily related to (III.1) by (III.2) or by (III.9)] exists at all on $\overline{M}$ and for which the required condition

$$\hat{g} < \hat{\eta} \tag{IV.1}$$

holds [this being the same as (II.2), but written now in terms of the conformal metrics, to emphasize that it applies to the *whole* of $\overline{M}$]. Here $\overline{M}$, with its given conformal metric $\hat{g}$, refers to the standard positive-mass exterior Schwarzschild solution [given by (III.1) in the range (III.3)] with any suitable interior [such as (III.4) in the range (III.5)—but the choice of interior turns out to be irrelevant] and with standard conformal boundary $\mathscr{I}^+ \cup \mathscr{I}^-$ [with the points of $\mathscr{I}^+$ given by (u, θ, ϕ) at $r = \infty$ and those of $\mathscr{I}^-$ given by (v, θ, ϕ) at $r = \infty$]. The topology of $\overline{M}$ is built up from the (Schwarzschild) space–time:

$$M = \text{int}\,\overline{M} \cong \mathbb{R}^4$$

and its conformal boundary $\partial\overline{M} = \mathscr{I}^+ \cup \mathscr{I}^-$:

$$\mathscr{I}^+ \cong \mathscr{I}^- \cong S^2 \times \mathbb{R}.$$

With the interior (III.4) in the range (III.5), M is actually an asymptotically simple space–time [3,5,6]. The standard notation $a \ll b$, for points $a, b \in \overline{M}$ is adopted [6,8] for the assertion "there exists a future-directed timelike curve in $\overline{M}$ from a to b." Also

$$I^+(a) = \{x \in \overline{M} \mid a \ll x\}, \qquad I^-(a) = \{x \in \overline{M} \mid x \ll a\}. \tag{IV.2}$$

I shall prove:

(IV.3) *Theorem* $\mathscr{I}^+ \subset I^+(a)$ *for each* $a \in \mathscr{I}^-$; *equivalently*, $\mathscr{I}^- \subset I^-(a)$ *for each* $b \in \mathscr{I}^+$.

An equivalent statement but which does not refer to $\mathscr{I}^\pm$ is:

(IV.4) *Theorem If* λ *and* μ *are endless timelike curves in* M, *then there exist points* $p \in \lambda$, $q \in \mu$ *with* $p \ll q$.

From these results can be derived, as a simple corollary, the required property

(IV.5) *Theorem There is no Minkowskian conformal metric* $\hat{\eta}$ *on* $\overline{M}$ *with* $\hat{g} < \hat{\eta}$.

Proofs Let us first establish the equivalence of (IV.3) with (IV.4). Note that another way of stating (IV.3) is

$$a \ll b, \text{ for every pair } a \in \mathscr{I}^-, b \in \mathscr{I}^+. \tag{IV.6}$$

Now, using the notion of causal boundary [3,7], we can interpret the point $a \in \mathscr{I}^-$ as an equivalence class of endless timelike curves having the same future in M (their points constituting the TIF in M representing a). Likewise, $b \in \mathscr{I}^+$ can be interpreted as an equivalence class of endless timelike curves with the same past in M (generating the TIP representing b). Let the endless timelike curves λ and μ represent a and b, respectively, so we can think of λ as acquiring a past endpoint a on $\mathscr{I}^-$ and μ as acquiring a future endpoint b on $\mathscr{I}^+$. The assertion $a \ll b$ amounts to saying that an endless timelike curve ν exists whose future agrees with that of λ and its past with that of μ. Now it is clear that *if* the statement of (IV.4) holds, then ν does indeed exist,

namely, consisting of the past-endless portion of λ up until the point p, the asserted timelike curve from p to q, and the future-endless portion of μ from q onward. (This jointed curve can be smoothed, if desired.) Conversely, suppose that λ and μ are as above and that ν has the same future as λ and the same past as μ. Choose a point w on ν. Now w must lie in the future of λ (since it lies in the future of ν), so a point p exists on λ with $p \ll w$. Similarly, q exists on μ with $w \ll q$. Hence $p \ll q$ holds, as required in (IV.4). There remains the possibility that one or both of the curves λ, μ in (IV.4) do not reach $\mathscr{I}^\pm$ at all, at their appropriate ideal endpoints but reach timelike infinity $i^\pm$ instead [3,5–7]. In these cases (IV.4) is again satisfied, but in a more trivial way.

Next, let us see why (IV.5) is a consequence of (IV.3) or (IV.4). Suppose a Minkowskian $\hat{\eta}$ exists on $\overline{M}$ with $\hat{g} < \hat{\eta}$. Now it is clear that (IV.4) would be false if "timelike" and "$\ll$" referred to the flat η metric. For example, in the usual T, X, Y, Z coordinates, with $d\sigma^2 = dT^2 - dX^2 - dY^2 - dZ^2$, the two branches of the η-timelike hyperbola $T^2 - X^2 + 1 = 0 = Y = Z$ are everywhere η-*spacelike* separated from one another. Suppose we let a be the past endpoint that one branch acquires on $\mathscr{I}^-$ and b the future endpoint that the other branch acquires on $\mathscr{I}^+$. By (IV.6), there is a g-timelike curve in $\overline{M}$ from a to b. But if $\hat{g} < \hat{\eta}$, this curve must also be η-timelike. But this is impossible as we have just seen. [The equivalence between (IV.4) and (IV.6) clearly holds equally for the $\hat{\eta}$-causality as for the $\hat{g}$-causality.] This contradiction establishes (IV.5) as a consequence of (IV.3) or (IV.4).

Finally, we must establish (IV.3). Let c be a point in M (with $r > r_0 > 3m$) and let γ be a null geodesic through c which is transverse to the source at c. Without loss of generality we can arrange c to have coordinates $(0, R, \pi/2, \pi/2)$ in the standard (t, r, θ, ϕ) coordinate system of (III.1) and, with the dot denoting $d/d\tau$, where τ is an affine parameter on γ,

$$\dot{r} = \dot{\theta} = 0 \qquad \text{at } t = 0. \tag{IV.7}$$

The standard geodesic equations for γ yield

$$\theta = \pi/2, \qquad r^2\dot{\phi} = A = \text{const}. \tag{IV.8}$$

and

$$[1 - (2m/r)]\dot{t} = B = \text{const}. \tag{IV.9}$$

We can normalize the affine parameter τ by choosing

$$B = 1 \quad \text{and} \quad \tau = 0 \quad \text{at } t = 0. \tag{IV.10}$$

Then, by (IV.7)–(IV.10) (and taking the positive sign for $\dot{\phi}$),

$$A = R[1 - (2m/R)]^{-1/2}$$

and [using (III.6): $u = t - r - 2m \log(r - 2m)$], we derive

$$u = -R - 2m \log(R - 2m) + \int_R^r f(\rho) d\rho$$

where

$$f(\rho) = \left(1 - \frac{2m}{\rho}\right)^{-1} \left\{ \left[1 - \frac{R^2}{\rho^2} \left(\frac{1 - 2m/\rho}{1 - 2m/R} \right) \right]^{-1/2} - 1 \right\}.$$

By inspection of (IV.13) we see that

$$\int_R^\infty f(\rho) d\rho \tag{IV.11}$$

converges—as it must, since γ reaches a point of $\mathscr{I}^+$ with finite u-value. We wish to examine how this u-value behaves as $R \to \infty$ and to show, in fact, that it tends to $-\infty$:

$$\lim_{R \to \infty} \left[-R - 2m \log(R - 2m) + \int_R^\infty f(\rho) d\rho \right] = -\infty. \tag{IV.12}$$

Now, to see this, we note

$$f(\rho) < \left(1 - \frac{2m}{R}\right)^{-1} \left\{ \left[1 - \frac{R^2}{\rho^2 (1 - 2m/R)} \right]^{-1/2} - 1 \right\} = g(\rho)$$

if

$$\rho \in (P, \infty) \qquad \text{where} \qquad P = R\left(1 - \frac{2m}{R}\right)^{-1/2} = R + O(1),$$

whence

$$\int_P^\infty f(\rho) d\rho < \int_P^\infty g(\rho) d\rho = R\left(1 - \frac{2m}{R}\right)^{-3/2} = R + O(1) \tag{IV.13}$$

by explicit integration. Furthermore we can estimate $\int_R^P f(\rho) d\rho$ using

$$f(\rho) = \left(1 - \frac{2m}{\rho}\right)^{-1}$$
$$\times \left(\left\{ \frac{(\rho - R)[\rho R(\rho + R) - 2m(\rho^2 + \rho R + R^2)]}{\rho^3 (R - 2m)} \right\}^{-1/2} - 1 \right)$$
$$< \left(1 - \frac{2m}{R}\right)^{-1} (\rho - R)^{-1/2} P^{3/2} (R - 2m)^{1/2} (2R^3 - 6mP^2)^{-1/2}.$$

provided that

$$\rho \in (R, P) \qquad \text{and} \qquad R > 5m$$

whence, using

$$\int_R^P (\rho - R)^{-1/2} d\rho = 2(P - R)^{1/2} = O(1),$$

we derive the fact that

$$\int_R^P f(\rho) d\rho = O(1) \tag{IV.14}$$

as $R \to \infty$. Combining (IV.14) with (IV.13), we obtain the result that (IV.11) is $R + O(1)$, so that substituting into the left-hand side of (IV.12), we derive the required result (IV.12), because of the presence of the logarithmic term.

This shows that whatever value of u is chosen, R can be made large enough that the null geodesic γ meets $\mathscr{I}^+$ at a u-value that is *less* than that chosen value. Moreover, because of the light-bending effect, γ will encounter an outgoing radial null geodesic β whose equation has the form

$$u = u_0 = \text{const}, \qquad \theta = \pi/2, \qquad \phi = \phi_0 = \text{const} \tag{IV.15}$$

for any value of ϕ_0 in the range

$$\phi_0 \in [\pi/2, \pi] \tag{IV.16}$$

(*including* the value $\phi_0 = \pi$), before reaching $\mathscr{I}^+$. Since u is an increasing function along γ, the value u_0 must be even less than the u-value attained at $\mathscr{I}^+$. We can likewise repeat the entire preceding argument in time-reversed form and attain the result that whatever value of v [given by (III.7)] is chosen, R can be made large enough that γ encounters, into the past, an incoming null geodesic α with equation

$$v = v_1 = \text{const}, \qquad \theta = \pi/2, \qquad \phi = \phi_1 = \text{const} \tag{IV.17}$$

for which v_1 is larger than the chosen v-value, and where ϕ_1 can take any value in the range

$$\phi_1 \in [0, \pi/2].$$

A jointed null geodesic, made up from pieces of α, γ, and β therefore connects the point $a \in \mathscr{I}^-$, with (v, θ, ϕ)-coordinates $(v_1, \pi/2, \phi_1)$, to the point $b \in \mathscr{I}^+$, with (u, θ, ϕ)-coordinates $(u_0, \pi/2, \phi_0)$. Thus [8]

$$a \ll b$$

as is required for (IV.6). (The jointed null geodesic can be smoothed, if desired, to yield to smooth timelike curve from a to b.) By a suitable rotation of the (θ, ϕ)-coordinate system, we can arrange for a to lie on any generator of $\mathscr{I}^-$ and b on any generator of $\mathscr{I}^+$. (The crucial case, in fact, is when the generators are opposite: $\phi_0 = \pi$, $\phi_1 = 0$.) And by allowing v_1 to be as large and positive as desired, and u_0 to be as large and negative as desired, we can cover all possibilities, thus establishing (IV.3). ■

V. Concluding Remarks

It should be clear from the preceding construction that the results of this paper are in no way specific to the Schwarzschild solution. One is concerned only with the nature of the space–time at large distances from the positive-mass source—evidently essentially with causal properties in the neighborhood of spacelike infinity i^0 [3,5]. Corresponding results are to be anticipated for *any* appropriately asymptotically flat space–time with positive mass.

One is tempted to use the fact that, whenever the null convergence condition holds [3,9] (a consequence of the weak energy condition and Einstein's equations) together with the genericity condition [3,9], every complete null geodesic in the space–time contains pairs of conjugate points [3,9]. This has the implication that for any point $a \in \mathscr{I}^-$ (assuming asymptotic simplicity), no generator of $\partial I^+(a)$ in the space–time extends all the way to $\mathscr{I}^+$. This imposes severe difficulties for the geometry of $\partial I^+(a) \cap \mathscr{I}^+$, if this set is to be nonvacuous, and appears to lead to a more general argument from which (IV.3) can be derived under much wider circumstances. The question seems also to be related to the positive energy conjecture [10,11] and to the details of the structure [12] of i^0. These matters will not be discussed here, as the result obtained in Section IV is adequate for the present purposes.

Acknowledgment

The author is grateful to the Miller Foundation and the University of California at Berkeley for their support and hospitality.

References

[1] Weinberg, S., "Gravitation and Cosmology." Wiley, New York, 1972.
[2] Duff, M. J., *in* "Quantum Gravity." C. J. Isham, R. Penrose, and D. W. Sciama, eds., p. 78–135. Oxford Univ. Press, London and New York, 1975.
[3] Hawking, S. W., and Ellis, G. F. R., "The Large Scale Structure of Spacetime." Cambridge Univ. Press, London and New York, 1973.
[4] Penrose, R., *Rev. Mod. Phys.* **37**, 215 (1965).
[5] Penrose, R., *Proc. R. Soc. London, Ser. A* **284**, 159 (1965).
[6] Penrose, R., *in* "Battelle Rencontres" (C. M. DeWitt and J. A. Wheeler, eds.), p. 171–189. Benjamin, New York, 1968.

[7] Geroch, R. P., Kronheimer, E. H., and Penrose, R., *Proc. R. Soc. London, Ser. A* **327**, 545 (1972).
[8] Penrose, R., "Techniques of Differential Topology in Relatively." *SIAM*, Philadelphia, Pennsylvania, 1972.
[9] Hawking, S. W., and Penrose, R., *Proc. R. Soc. London, Ser. A* **314**, 529 (1970).
[10] Geroch, R. P., *Ann. N.Y. Acad. Sci.* **224**, 125 (1973).
[11] Schoen, R., and Yau, S.-T., *Commun. Math. Phys.* **65**, 45 (1979).
[12] Ashtekar, A., and Hansen, R. O., *J. Math. Phys.* **19**, 1542 (1978).

Gen Relativ Gravit (2010) 42:641–646
DOI 10.1007/s10714-009-0912-9

RESEARCH ARTICLE

Gravity from self-interaction redux

Stanley Deser

Received: 15 October 2009 / Accepted: 1 November 2009 / Published online: 18 November 2009

Abstract I correct some recent misunderstandings about, and amplify some details of, an old explicit non-geometrical derivation of GR.

Keywords Graviton self-coupling · Stress-tensors · Spin 2 sources

Long ago [1], I presented a compact derivation of GR from an initial free flat space long-range symmetric spin two field: Since special relativity replaces the matter Newtonian scalar mass density by its stress-tensor, a tensor must likewise replace the scalar "potential". Consistency then forces this field to couple to its its own stress tensor if it is to allow any matter coupling: it either stays free- and dull- or its stress-tensor must be added to that of matter as the field's source. This bootstrap was then explicitly performed in GR by exploiting its first derivative, cubic, $L \sim p\dot{q} - qp^2$, rather than its more familiar second-order non-polynomial $L(q)$, form. The process was also illustrated in the simpler, but precisely analogous, context of deriving (non-linear)YM from a multiplet of free Maxwell fields, which must likewise self-couple to accept non-abelian sources. Subsequently, two extensions of [1] were found: First, it was generalized to allow starting from any constant curvature background, where spin 2 is consistently defined [2]. The cosmological term could then also be included in the bootstrap. Second, a tree-level quantum derivation [3] (later generalized to include SUGRA [4]) provided an alternate framework, where the irrelevance of inherent field redefinition ambiguities and freedoms is particularly clear.

S. Deser (✉)
Physics Department, Brandeis University, Waltham, MA 02454, USA
e-mail: deser@brandeis.edu

S. Deser
Lauritsen Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

Recently, however, there have appeared lengthy, (if not mutually consistent) critiques [5,6] of [1]. This note addresses and resolves their concerns, both conceptual and technical, by expanding on the, perhaps too concise, original. For orientation, we start with the list of main worries and the short answers.

1. The self-coupling idea, while appealing, does not work out concretely; also,the gravitational stress-tensor is ill-defined.

These worries stem from too narrow a view of self-coupling and a too broad one of non-uniqueness. Self-coupling means that the right-hand side of the original free-field equations, in one of its possible incarnations, acquires as a source the field's own total stress tensor. This will be (re-)derived below, using the equivalent but more convenient Ricci, rather than Einstein, form of the equations. A related complaint was that the coupling did not appear in the naive, $h^{\mu\nu}T_{\mu\nu}$, form in the action. True, but irrelevant: to repeat, the only physical requirement is that, in the field equations, the full $T_{\mu\nu}$ become the source of the originally free field; the action's sole job is to yield these, and it does—see (13) below. Non-uniqueness of the stress tensor: it is indeed always undetermined up to identically conserved super-potentials. Further, while the one place where this non-uniqueness is relevant, namely when the stress tensors become local sources, is here, it is also precisely here that all such ambiguities can be absorbed, as we shall see, by harmless field redefinitions. Another non-uniqueness pseudo-problem is that free gauge fields of spin >1 cannot possess (abelian) gauge-invariant stress tensors; this truism actually turns out to be a plus: only full GR recaptures the initial invariance, but now in non-abelian form, at the (satisfactory!) price of forfeiting any physical significance for its own stress-tensors, a fact also known as the equivalence principle. The only restriction on the initial stress-tensor(s) is that they be symmetric so they can drive the graviton's symmetric field equations; further, only they can define angular momentum.

2. The GR action's non-analytic dependence on the Einstein constant κ cannot be obtained perturbatively starting from the, $\sim \kappa^0$, free field.

This worry will be easily dispatched in its place; simply, the final $1/\kappa^2$ dependence arises from a constant field rescaling of the (analytic) result to connect the field theoretical and geometrical variables' dimensions.

3. The theory's second derivative order was an assumption.

This is as true here as it was for Einstein and Newton! Formally, GR is but one of an infinite set of geometrical models, with as as many derivatives as desired (e.g., $L \sim RD^nR$) ... Observation determines the initial kinematics, excluding (to leading order at least) scalar-tensor mixtures and higher derivative terms. Most relevant for us, second derivative order together with infinite range (any finite range makes qualitatively wrong weak-field predictions [7–12]) means that a gauge invariant (i.e., ghost-free) massless tensor field is the initial, special relativistic, mediator of matter–matter forces (their attractive sign then being a built-in bonus [13]).

4. Total divergences and surface terms are important.

Yes, but not to obtain Euler–Lagrange equations from an action. Surface terms are indeed physically useful in GR, but not because of their presence in its action, contrary to myth.

5. As (correctly) noted in [5], there have many other attempts at deriving GR from self-coupling, none of which succeeded: their approach being purely metric, the infinite summations needed to reach non-polynomial metric GR have never been performed. Instead, they were replaced by such statements as "what else could it sum to?" and "the sum must be general covariant, ergo GR".

Agreed. In particular the covariance of the final result, in the strong sense of being achieved without involving an external metric, does emerge here without being postulated; likewise, "summation" is trivial.

For maximum clarity, we focus on the logic, with a minimum of formalism and indices; that can be found in [1]. The flat space, first order, Fierz–Pauli massless spin 2 Lagrangian is

$$L_2 = h^{\mu\nu}(\partial_\alpha \Gamma^\alpha_{\mu\nu} - \partial_\mu \Gamma^\alpha_{\alpha\nu}) + \eta^{\mu\nu}(\Gamma^\alpha_{\mu\nu}\Gamma^\beta_{\beta\alpha} - \Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu}) \tag{1}$$

The two independent variables are the Minkowski tensors $(h^{\mu\nu}, \Gamma^\alpha_{\mu\nu})$, with dimension (L^{-1}, L^{-2}) as befits their "(q, p)" nature; η is the Minkowski metric. The resulting first order field equations

$$\partial_\alpha \Gamma^\alpha_{\mu\nu} - \frac{1}{2}(\partial_\mu \Gamma^\alpha_{\alpha\nu} + \partial_\nu \Gamma^\alpha_{\alpha\mu}) = 0 \tag{2}$$

$$\partial_\alpha h^{\mu\nu} - \partial_\mu h^{\nu\alpha} - \frac{1}{2}\eta_{\mu\nu}\partial_\alpha h^\beta{}_\beta = 2\Gamma^\alpha_{\mu\nu} - \eta^\alpha_\mu \Gamma^\beta_{\beta\nu} - \eta^\alpha_\nu \Gamma^\beta_{\beta\mu} \tag{3}$$

are equivalent to

$$2R^L_{\mu\nu}(h) \equiv \partial_\beta \partial^\beta \left(h^{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h^\alpha_\alpha\right) - \partial_\nu \partial_\alpha h^{\mu\alpha} - \partial_\mu \partial_\alpha h^{\nu\alpha} \tag{4}$$

in terms of the linearized Ricci (rather than Einstein) tensor.[1] [Our $h^{\mu\nu}$ is related to the usual covariant metric deviation $h_{\mu\nu}$ by $h^{\mu\nu} = -h_{\mu\nu} + (1/2)\eta_{\mu\nu}(h_{\alpha\beta}\eta^{\alpha\beta})$]. Note however that our $h^{\mu\nu}$ is NOT the start of an expansion, but is the total deviation, from its Minkowski value, of the full contravariant metric density.

The full GR, Palatini, Lagrangian we want to derive is

$$\begin{aligned} L_{EH}(\mathcal{G}, \Gamma) &= \kappa^{-2}\mathcal{G}^{\mu\nu}R_{\mu\nu}(\Gamma) \\ &= \kappa^{-2}\mathcal{G}^{\mu\nu}(\partial_\alpha \Gamma^\alpha_{\mu\nu} - \partial_\mu \Gamma^\alpha_{\alpha\nu} + \Gamma^\alpha_{\mu\nu}\Gamma^\beta_{\beta\alpha} - \Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu}); \end{aligned} \tag{5}$$

$\mathcal{G}$ is the contravariant metric density, Γ the (independent) affinity. The chief differences between (1) and (5) are that there is no background space dependence in (5), and that

[1] For comparison, the first order vector theory equivalents are the initial, $L_1 \sim F curl A - F^2$ and $L_{YM} \sim L_1 + gFAA \sim p\dot{q} - p^2 + pq^2$ as final, forms; they are spelled out in [1].

it is cubic (rather than quadratic) in the fields. This latter property is its compelling attraction for us, in contrast to the second order metric formulation's non-polynomial dependence on both the metric and its inverse through the affinity's metric dependence. The GR equations, from varying $\mathcal{G}$ and Γ independently, are

$$R_{\mu\nu}(\Gamma) \equiv \partial_\alpha \Gamma^\alpha_{\mu\nu} - \frac{1}{2}\partial_\mu \Gamma^\alpha_{\alpha\nu} - \frac{1}{2}\partial_\nu \Gamma^\alpha_{\alpha\mu} + (\Gamma^\alpha_{\mu\nu}\Gamma^\beta_{\beta\alpha} - \Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu}) = 0, \tag{6}$$

$$-\partial_\alpha \mathcal{G}^{\mu\nu} + \mathcal{G}^{\mu\nu}\Gamma^\lambda_{\lambda\alpha} - \mathcal{G}^{\mu\rho}\Gamma^\nu_{\alpha\rho} - \mathcal{G}^{\nu\rho}\Gamma^\mu_{\alpha\rho} = 0, \tag{7}$$

and reduce to $R_{\mu\nu}(\mathcal{G}) = 0$ upon inserting $\Gamma(\mathcal{G}) \sim \mathcal{G}^{-1}\partial\mathcal{G}$ into (6). Note that the geometrical variables' dimensions are ($\mathcal{G} \sim L^0, \Gamma \sim L^{-1}$). We will see that the non-analyticity of (5) is purely apparent, being removable by constant rescalings. It is useful for the sequel to express this desired answer in flat space notation by expanding (5) in terms of $\mathcal{G} = \eta + \kappa h$ (κ restores h's original dimension L^{-1}) and to restore its old dimension to Γ, by defining $\overline{\Gamma} = \kappa^{-1}\Gamma$; we now drop all indices to concentrate on the form and logic:

$$L_{EH}(h, \overline{\Gamma}) = \kappa^{-1}\eta\partial\overline{\Gamma} + (h\partial\overline{\Gamma} + \eta\overline{\Gamma}\,\overline{\Gamma}) + \kappa h\overline{\Gamma}\,\overline{\Gamma}. \tag{8}$$

The first term being an irrelevant total divergence, κ now appears quite tamely in the rest of (8), disposing nicely of that worry. The middle terms are precisely the quadratic free field Lagrangian (1). The cubic term, $\kappa h\overline{\Gamma\Gamma} \equiv \kappa h S$ is of course supposed to supply the heralded self-coupling of h to its stress tensor in the field equations (as we will check it does), the very reason S is not itself the stress tensor. Given this flat space form of GR, it remains to show that the cubic term in (8) is the right choice: does it provide just the right (whatever that is) stress tensor source of the free field—middle terms'—field equation? The justification has three parts: first obtaining the stress tensor(s) of the middle terms' action, then showing why its non-uniqueness (including abelian gauge-variance) is harmless, and finally verifying that the chosen cubic term (the one that agrees with L_{EH}) indeed produces this stress tensor.

First, the stress tensor: we use the Belinfante prescription: write the flat space action $\mathcal{I}$ covariantly with respect to a fictitious auxiliary metric (for us a contravariant density) $\gamma^{\mu\nu}$, vary the resulting action with respect to it, then set it back to $\eta^{\mu\nu}$ in the resulting variation. The result is a symmetric on-shell, trace-shifted, stress tensor. In (1), there are two places to covariantize: the obvious $\eta\Gamma\Gamma \to \gamma\Gamma\Gamma$ and $h\partial\Gamma \to hD(\gamma)\Gamma$, where D is the covariant tensor derivative involving the auxiliary Christoffel symbols $\sim$ $(\partial\gamma)$ to first order. Manifestly,

$$\overline{T}_{\mu\nu} \equiv T_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\,\mathrm{tr}\,T \equiv (\delta\mathcal{I}/\delta\gamma)\,|_{\gamma=\eta} \sim \partial(h\Gamma) + \Gamma\Gamma. \tag{9}$$

Next (non-)uniqueness: to the Belinfante tensor (of any system) may be added any identically conserved super-potential

$$\Delta_{\mu\nu} = \partial^\alpha\partial^\beta H_{[\mu\alpha][\nu\beta]} = \Delta_{\nu\mu}, \quad \partial_\mu\Delta^{\mu\nu} \equiv 0, \tag{10}$$

where H is any 4-index function with the symmetries of the Riemann tensor (to keep Δ symmetric). [These contributions may also be thought of as the result of adding non-minimal couplings $\sim R_{\alpha\beta\gamma\delta}(\gamma)F^{\alpha\beta\gamma\delta}h, \Gamma$ to the original action (before varying γ)]. But identical conservation of Δ means precisely that it can be absorbed by field redefinition: the usual linearized Einstein equation is of the form

$$G^L_{\mu\nu}(h) = \mathcal{O}_{\mu\nu\alpha\beta}h^{\alpha\beta}, \quad \partial^\mu \mathcal{O}_{\mu\nu\alpha\beta} \equiv 0. \tag{11}$$

Hence any identically conserved source can simply be removed by a corresponding shift in h. [The initial Belinfante part, not being a super-potential, cannot be shifted away]. Finally, we must show that the cubic term in (8) indeed yields the desired field equation, with the stress tensor (9) as source of the free field. That is, we want to verify that the full field equation reads $R^L_{\mu\nu}(\Gamma(h)) \sim \kappa\overline{T}_{\mu\nu}$. The Einstein equations (6,7) are, dropping the overbars and expanding $\mathcal{G}$,

$$\partial\Gamma + \kappa\Gamma\Gamma = 0, \quad \Gamma = \partial h + \kappa h\Gamma. \tag{12}$$

Differentiating the second and inserting it into the first equation gives precisely the promised second order form

$$\partial^2 h = \kappa[\partial(h\Gamma) + \Gamma\Gamma] \equiv \kappa\bar{T}. \tag{13}$$

More explicitly, the left side is $R^L_{\mu\nu}(\Gamma(h))$, while the right is just the $\overline{T}_{\mu\nu}$ of (9) if (and only if) we use the cubic term of the GR action (5). Equally important, the bootstrap stops here because this cubic term in the action does not generate any further (cubic) stress-tensor correction, being both η-and derivative-independent. This completes our exegesis.

Sources: it is rather obvious that any matter action must couple to the final GR through its variables ($\mathcal{G}$, Γ) or $\mathcal{G}$ alone, and do so covariantly in order to respect the GR equation's Bianchi identities by having an (on-shell) covariantly conserved metric variation. But this is just Noether's theorem: any system's stress-tensor, namely the variation of its action with respect to the metric that makes it invariant, is covariantly conserved by virtue of its own field equations, irrespective of the equations (if any), satisfied by the metric.

In summary, I have annotated the steps involved in the non-geometric derivation [1] of GR from special relativistic field theory as the unique consistent self-interacting system (13) extending the initial free massless spin 2. The main ingredients were: computing the field's standard Belinfante stress tensor, invoking field-redefinition freedom to neutralize its non-uniqueness, performing a constant field rescaling to relate geometric and field theoretic variables, and (most important) employing the cubic, Palatini, first order forms to permit explicit, trivial, summation. It goes without saying that this non-geometrical interpretation of GR, far from replacing Einstein's original geometrical vision, is a tribute to its scope.

Acknowledgments This work was supported by NSF grant PHY 07-57190 and DOE grant DE-FG02-92ER40701.

References

1. Deser, S.: Gen. Relativ. Gravit. **1**, 9 (1970), reprinted as gr-qc/0411023
2. Deser, S.: Class Quantum Grav. **4**, L 99 (1987)
3. Boulware, D., Deser, S.: Ann. Phys. **89**, 193 (1975)
4. Boulware, D., Deser, S., Kay, J.: Physica **96A**, 141 (1979)
5. Padmanabhan, T.: Int. J. Mod. Phys. D **17**, 367 (2008), gr-qc/0409089
6. Butcher, L., Hobson, M., Lasenby, A.: Phys. Rev. D **80**, 084014 (2009), gr-qc/0906.0926
7. van Dam, H., Veltman, M.: Nucl. Phys. B **22**, 397 (1970)
8. Zakharov, V.: JETP Lett. **12**, 312 (1970)
9. Faddeev, L., Slavnov, A.: Theor. Math. Phys. **3**, 18 (1970)
10. Wong, S.: Phys. Rev. D **3**, 945 (1971)
11. Kogan, I., Mouslopoulos, S., Papazoglou, A.: Phys. Lett. B **503**, 173 (2001), hep-th/0011138
12. Porrati, M.: Phys. Lett. B **498**, 92 (2001), hep-th/0011152
13. Deser, S.: Am. J. Phys. **73**, 6 (2005), gr-qc/0411026