PHYSICAL REVIEW LETTERS

VOLUME 50 | 7 MARCH 1983 | NUMBER 10

Causality Versus Gauge Invariance in Quantum Gravity and Supergravity

Claudio Teitelboim
Center for Theoretical Physics, The University of Texas at Austin, Austin, Texas 78712
(Received 6 December 1982)

It is proposed that one should admit in the path integral for the quantized gravitational field only those space-times for which the final three-geometry is located in the future of the initial one. As a consequence, and unlike the situation for the Yang-Mills field, the resulting causal amplitude is not annihilated by all the gauge (surface deformation) generators. In supergravity the causal amplitude turns out not to be annihilated by the local supersymmetry generators either.

PACS numbers: 04.60.+n, 11.15.+q, 11.30.Pb

At present, a satisfactory quantum theory of gravitation does not exist. However, one can nevertheless ascertain some general features which, one believes, will be incorporated in the yet to be found complete scheme.

One such feature is the fact that space-time understood as a pseudo-Riemannian manifold is a purely classical concept and, as such, disappears upon passing to the quantum theory. Indeed, as forcefully emphasized by Wheeler,[1] space-time is the classical history of three-dimensional space and it is in this respect the analog of the world line of a particle.

Since there is no space-time when the gravitational field is quantized there is in particular no notion of time, and it would seem that there is no place either for the concept of causality.

However, a closer analysis reveals that in spite of the above remarks one can and, we believe, must, incorporate the notion of causality at a basic level in the quantum theory of the gravitational field. Furthermore, as we will see below, taking causality into account has definite consequences, among them one which is perhaps surprising, namely, that in a precise sense an important part of the gauge freedom of the classical theory does not persist in the quantum theory. In this respect the quantum theory of gravitation appears to be quite different from that of the Yang-Mills field.

The key point in the analysis is to recall that, as revealed by the path-integral formulation of quantum mechanics, the reason for the fading away of the classical history is the fact that the quantum mechanical propagation amplitude is "made" out of many histories and not just out of the classical one. From this point of view a history is a more elementary concept than that of a quantum mechanical amplitude.

This observation suggests that the notion of causality should be incorporated at the level of the individual histories prior to folding those histories into the quantum amplitude by means of the path integral. The reason for this proposal is the fact that each individual history has a pseudo-Riemannian structure and, with it, a light cone and a notion of past and future.

To proceed with the analysis it is necessary to write down the action integral for the gravitational field. In doing so we will assume for simplicity that the three-dimensional space is compact. The analysis and conclusions remain, however, valid for the open asymptotically flat case, on which some special comments will be made at the end.

The gravitational field action in Hamiltonian form reads,

$$S = \int dt\,dx(\pi^{ik}\dot{g}_{ik} - N^{\mu}\mathcal{H}_{\mu}). \qquad (1)$$

Here $g_{ik}(x, t)$ is the metric of the spacelike surface of constant t and π^{ik} is its canonically conjugate momentum. The functions $N^{\mu} = (N^{\perp}, N^{i})$ describe the deformation that connects the surface of time t with that of time $t + \delta t$. Here N^{i} is the usual "shift vector," whereas $N^{\perp}$ is equal to $[\det(g_{ij})]^{-1/2}$ times the usual "lapse function." The surface deformation generators $\mathcal{H}_{\mu} = (\mathcal{H}_{\perp}, \mathcal{H}_{i})$ are constructed from the g_{ik} and the π^{ik}.

A fundamental property of the action (1) is its invariance under the transformation defined by

$$\delta F = [F, \int \epsilon^{\mu}\mathcal{H}_{\mu}dx], \qquad (2)$$

where F is either g_{ij} or π^{ij}, together with an accompanying transformation law for N^{μ}. (See Ref. 2 and bibliography therein.) Such a transformation is called a gauge transformation because the ϵ^{μ} are arbitrary functions of space and time. It may be thought of as the Hamiltonian version of an infinitesimal change of space-time coordinates, induced by a vector field with normal and tangential components $\epsilon^{\perp}$, ϵ^{i}.

From (2) one learns that the $\mathcal{H}_{\mu}$ are the generators of gauge transformations. They also play the role of a "many-time" Hamiltonian since $\int N^{\mu}\mathcal{H}_{\mu}$ is the Hamiltonian in (1). This is not surprising since in a generally covariant theory both time evolution and gauge transformations correspond to localized space-time displacements.

Einstein's equations are the conditions for the action (1) to have an extremum under variations of g_{ij}, π^{ij}, and N^{μ}, for which the spatial metric g_{ij} is fixed at t_1 and t_2 up to a change of coordinates. In other words, what is kept fixed in the action principle are the initial and final three-geometries G_1, G_2. Upon passing to the quantum theory G_1 and G_2 become the arguments of the propagation amplitude, which is obtained by summing the exponential of i times the action (1) over histories (space-times) that interpolate between both three-geometries.

The amplitude turns out to have the simple representation[2]

$$K[G_2, G_1]$$
$$= \int D[f]D[T^{\perp}]\langle 2|\exp[-i\int T^{\perp}\mathcal{H}_{\perp}{}^{\mathrm{eff}}dx]|f(1)\rangle. \qquad (3)$$

The meaning of the quantities appearing in (3) is the following: The generator $\mathcal{H}_{\perp}{}^{\mathrm{eff}}$ differs from the classical one appearing in (1) by the addition of a ghost contribution $\mathcal{H}_{\perp}{}^{\mathrm{ghost}}$ of purely quantum mechanical nature. The states $|1\rangle$ and $|2\rangle$ are eigenstates of the metric field g_{ij} with eigenvalues $g_{ij}(1)$ and $g_{ij}(2)$, respectively, and of the anticommuting ghost fields which enter into $\mathcal{H}_{\perp}{}^{\mathrm{ghost}}$ with eigenvalue zero. The symbol $f(1)$ means that $g_{ij}(1)$ undergoes a change of spatial coordinates by the diffeomorphism $x \to f(x)$ and the notation $D[f]$ represents the invariant measure over the diffeomorphism group. The measure $D[T^{\perp}]$ is given by the infinite product of $dT^{\perp}(x)/T^{\perp}(x)$ over all points of space.

The function $T^{\perp}(x)$ is a measure of the total pointwise proper time separation between the initial and final three-geometries. More precisely, for each history contributing to the path integral one slices through the region between G_1 and G_2 by a family of intermediate surfaces determined by the "proper time gauge conditions," $\dot{N}^{\perp} = 0$, $N^{i} = 0$, a set of relations which also fix the spatial coordinate system throughout relative to that on the initial surface. With this construction one has $T^{\perp}(x) = (t_2 - t_1)N^{\perp}(x)$.

The path integration is first performed over all histories with a given $T^{\perp}(x)$ and with a fixed coordinate system on the final surface. This yields the matrix element in the integrand of (3) prior to the action of f on $g_{ij}(1)$. Then one sums over all (permissible) locations of the final surface relative to the initial one, which gives the integral over $T^{\perp}(x)$ in (3). Finally one averages over all possible choices of the coordinate system on the final surface relative to that on the initial one. This gives the integration over f in (3).

The representation (3) permits one to implement the requirement of causality by demanding that not all histories having G_1 and G_2 as its boundaries should contribute to the path integral. Rather, one demands that *only those histories for which G_2 lies in the future of G_1 should be admitted.* This restriction may be incorporated directly into the amplitude (3) by restricting the range of integration over $T^{\perp}(x)$, *to include positive proper times only.* Note that one requires that the final surface be *wholly* in the future of the initial one [$T^{\perp}(x)$ positive for *every* x]. Thus, in particular, the two surfaces are not allowed to intersect each other.

Now, consider the action of the generator $\mathcal{H}_{i}$ of tangential deformations (spatial reparametrizations), understood as quantum field operator, on the amplitude $K_+[G_2, G_1]$ obtained by only including positive proper times in (3). That is, let $\mathcal{H}_i$ act on $K_+[G_2, G_1]$ considered as a wave functional in $g_{ij}(1)$, for fixed $g_{ij}(2)$ [the same analysis may be performed on $g_{ij}(2)$]. One obtains

$$\mathcal{H}_i(x)(1)K_+[G_2, G_1] = 0. \qquad (4)$$

Equation (4) follows from the fact that the group average over changes of the spatial coordinates renders K_+ invariant under reparametrizations in both $g_{ij}(1)$ and $g_{ij}(2)$. (Thus the notation $K_+[G_2, G_1]$ is justified.) On account of (4) one says that the quantum mechanical amplitude is invariant under tangential deformations.

However, the situation with respect to $\mathcal{H}_\perp$ is different. Indeed, one finds

$$\mathcal{H}_\perp(x)(1)K_+[G_2, G_1] \neq 0, \qquad (5)$$

which shows that the amplitude is not invariant under normal deformations.

The reason for the lack of equality in (5) is the restriction to positive $T^\perp(x)$ in (3). That restriction means that, as required by the principle of causality, one integrates only over half of the possible locations of the final surface relative to the initial one. Or, in group-theoretical language, one averages the amplitude over only half of the space of possible deformations of the initial surface. This incomplete average makes the result not invariant under the action of the corresponding generator $\mathcal{H}_\perp$.

It is important to emphasize that it is the domain of integration over $T^\perp$ in (3) rather than the measure of integration which is responsible for (5). Indeed one may define a different noncausal, amplitude by extending the range of integration in (3) to cover all possible normal deformations, $-\infty < T^\perp < +\infty$. That amplitude, which will be denoted by Δ, may be shown to satisfy

$$\mathcal{H}_\mu(x)(1)\Delta[G_2, G_1] = 0 \qquad (6)$$

for all four values of μ.

The preceding discussion has a close analog in the quantum mechanics of a relativistic scalar particle in an external field (positron theory,[3] with the spin degrees of freedom omitted). In the parallel, the analog of the three-geometry G is the space-time position x^μ. The space coordinates x of the gravitational problem have no counterpart since the particle has zero space dimensions. So there is only one gauge generator which takes the form $\mathcal{H} = p^2 + m^2 + V(x)$. The equation corresponding to (3) is the proper-time representation which expresses the Feynman propagator $K_+(x_2, x_1)$ as an integral from $T=0$ to $T=+\infty$ of the matrix element $\langle x_2|\exp(-iT\mathcal{H})|x_1\rangle$. That propagator obeys

$$\mathcal{H}(1)K_+(x_2, x_1) = \delta(x_2, x_1) \neq 0. \qquad (7)$$

On the other hand if in that same representation one extends the integration from $T=-\infty$ to $T=+\infty$, one obtains a function $\Delta(x_2, x_1)$ which is proportional to $\langle x_2|\delta(\mathcal{H})|x_1\rangle$ and hence satisfies

$$\mathcal{H}(1)\Delta(x_2, x_1) = 0. \qquad (8)$$

Equation (5) has a serious consequence for the interpretation of the theory, especially in the case of asymptotically flat space. [The previous discussion is valid for asymptotically flat space provided the integration in (3) is performed over functions $T^\perp$ with fixed asymptotic behavior $T^\perp \to T_\infty{}^\perp$, for $r \to \infty$, and for diffeomorphisms f which become a fixed element of the rotation group at infinity. Equations (4)–(6) are understood to be weighted with a testing function that vanishes asymptotically.]

Indeed, when the space is asymptotically flat, the point of view is usually adopted that the gravitational field may be treated along the same lines as the Yang-Mills field. The group of changes of coordinates in four dimensions which become the identity at infinity replaces then the product of the internal Lie group over all points of space-time. Physically speaking this means that for an observer at infinity it is irrelevant how the flat hypersurface of constant Minkowskian time on which he stands is continued "inside."

However, this point of view is only tenable if the transition amplitude itself is insensitive to the continuation or, what is the same, if it is gauge invariant. In Hamiltonian langauge this means that the amplitude must be annihilated by all the gauge generators, as indeed happens in Yang-Mills theory.

It follows that, in asymptotically flat space, the gravitational field could be treated as the Yang-Mills one if the propagation amplitude were taken to be the $\Delta[G_2, G_1]$ defined above, which obeys (6). On the other hand the approach is not tenable if one adopts the causal amplitude K_+.

In the latter case the amplitude is sensitive to the way in which the surface is continued inside, because one can only perform the continuation of the final surface without wandering into the past of the initial one, and vice versa. This limitation disrupts the group structure of the four-dimensional diffeomorphisms and makes K_+ not gauge invariant.

It should be emphasized here that the possibility of naturally splitting the space-time diffeomorphism group into two halves (deformations to the past and deformations to the future) can only be achieved with the help of the pseudo-Riemannian metric—which is used to select the normal direction needed to define $N^\perp$, $T^\perp$—and is not feasible at the topological level. A similar option does not exist for an internal symmetry group or, for that matter, for the purely tangential deformations (changes of spatial coordinates). Therefore the only amplitude one can reasonably define in Yang-Mills theory is annihilated by all the gauge generators. By the same reason, in the gravitational case, both Δ and K_+ are annihilated by $\mathcal{H}_i$.

One may extend the above analysis to supergravity, whose analog in the sense of Eqs. (7) and (8) is the theory of a spin-$\frac{1}{2}$ particle. In supergravity there appear in the exponent of the amplitude (3) extra terms of the form $\int \theta^A S_A dx$ where S_A are the generators of localized supersymmetry transformations and $\theta^A(x)$ are the anticommuting analog[4] of the "proper time" $T^\perp(x)$. The causality condition $T^\perp \geqslant 0$ (there is no causality condition on θ^A) then implies not only (5) but also

$$S_A(x)K_+ \neq 0, \tag{9}$$

as may be directly seen from (4) and (5) and the basic anticommutation rule,[5]

$$[S_A(x), S_B(x')]_+ = (\gamma^0\gamma^\mu)_{AB}\mathcal{H}_\mu \delta(x, x'). \tag{10}$$

On the other hand, as (10) also shows, the amplitude Δ for supergravity obeys

$$S_A(x)\Delta = 0, \tag{11}$$

in addition to (6). Thus in supergravity the causal amplitude K_+ is not invariant under localized supersymmetry transformations whereas the noncausal amplitude is Δ supersymmetry invariant.

Therefore, it appears that in both gravity and supergravity one is faced with the alternative of preserving either gauge invariance or causality. It is the opinion of this author that one should preserve causality. In the case of positron theory[3] this turns out to be the correct choice ultimately because only by using the Feynman propagator does one obtain a unitary amplitude. (If one replaces K_+ by Δ in, say, a perturbation scheme, the resulting amplitude is not unitary.) Whether or not a similar situation will arise for the quantized gravitational field remains to be seen.

This work was supported in part by National Science Foundation under Grant No. PHY-8216715. The author is a recipient of a John S. Guggenheim Fellowship.

[1] J. A. Wheeler, in *Analytical Methods in Mathematical Physics*, edited by R. Gilbert and R. Newton (Gordon and Breach, New York, 1970).

[2] C. Teitelboim, Phys. Rev. D 25, 3159 (1982), and "The Proper Time Gauge in Quantum Theory of Gravitation," Phys. Rev. D, to be published, and "Quantum Mechanics of the Gravitational Field in Asymptotically Flat Space," Phys. Rev. D, to be published.

[3] R. P. Feynman, Phys. Rev. 76, 749 (1949); see also M. Henneaux and C. Teitelboim, Ann. Phys. (N.Y.) 143, 127 (1982).

[4] C. Teitelboim, in *Superspace and Supergravity*, edited by S. W. Hawking and M. Rocek (Cambridge, Univ. Press, Cambridge, England, 1981), pp. 185–194.

[5] C. Teitelboim, Phys. Rev. Lett. 38, 1106 (1977); R. Tabensky and C. Teitelboim, Phys. Lett. 69B, 240 (1977).