Further Discussion of the Role of Electromagnetic Potentials in the Quantum Theory

Y. AHARONOV*
Yeshiva University, New York, New York

AND

D. BOHM
Birkbeck College, University of London, London, England

(Received October 31, 1962)

In this paper, we present a further discussion of the role of electromagnetic potentials in the quantum theory, aimed at clarifying some of the points made in earlier papers, and indicating some extensions of these earlier ideas. In addition, we go into the problem of subsidiary conditions and questions concerning the full observability of the potentials. In the course of our discussion, we answer certain recent objections of Belinfante and De Witt.

1. INTRODUCTION

SOME time ago, a number of papers[1–3] were published, bringing out the fact that electromagnetic potentials have a new kind of significance in quantum theory, which does not appear in classical physics. Since then, a brief paper was published by De Witt,[4] in which he criticized one of the points that we had made, viz., that the potentials play an essential part in expressing the fact that the electron-electromagnetic interaction is a *local* one (i.e., that the interaction is nonzero only at the points where there are charges). In an accompanying paper[5] we answered these criticisms. Since then, however, a paper by Belinfante[6] has appeared, in which he objects to certain points raised in our answer to De Witt.

After carefully considering Belinfante's criticisms, and reconsidering De Witt's note, it has become clear to us that a systematic restatement of our position on this problem is now desirable, because what we wished to say has apparently not been sufficiently clearly understood. In addition, our ideas on this subject have meanwhile been developing in several directions, and we feel that an indication of the general line of these developments will also help to clarify some of the problems involved. In this article, we shall, therefore, try to present such a restatement of our general position along with a discussion of some further developments in our ideas concerning the role of electromagnetic potentials in quantum theory. In the course of the paper, we shall answer Belinfante's objections, and we shall also supplement our earlier reply to De Witt's note.

2. ON THE PRINCIPLE OF LOCALIZABILITY AND ITS RELATION WITH THE ELECTROMAGNETIC POTENTIALS

The main point that we wished to stress in our earlier paper[5] was that when quantum electrodynamics is expressed in terms of the potentials, one obtains both a complete set of local commuting "observables" *and* a set of purely local equations of motion. It was certainly not our intention, of course, to insist that a correct theory *must* be local in this sense. Indeed, we explicitly pointed out that physics is currently in a state of flux, such that there are even good reasons to suppose that this requirement of locality must somehow eventually either be changed or given up altogether. However, we did wish to stress that this *principle of localizability* (which may be called "physical-geometrical") is, *in fact*, now playing an important part in helping to define the *mathematical form* of current theories. We also wished to point out that there did not exist, to our knowledge, a correspondingly clear and natural way of expressing this principle of localizability in terms of the field strengths alone, without the use of potentials. It was in this sense that we meant to state that the potentials are playing an "essential" part in giving a mathematical expression to certain physically (and, of course, geometrically) significant features of current quantum electrodynamical theories.

Now, De Witt[4] and Belinfante[6] objected to the content of the above statements, and supported their objections by producing, at least in a formal sense, some examples of explicit theories based only on a local set of gauge-invariant "field observables," not requiring the introduction of the potentials. Thus, De Witt began by starting with the usual theory, and making a certain nonlocal gauge transformation, $\psi' = e^{iS}\psi$, where S is an integral of the field observables over a certain somewhat arbitrarily chosen path. In terms of ψ', he then obtained a set of field equations, very similar to the usual ones, except that where the vector potential usually enters, there now appears a nonlocal integral

$$A_\mu'(x^\alpha) = \int_{-\infty}^{0} F_{\nu\mu}(z^\lambda) \frac{\partial z^\nu}{\partial \xi} d\xi. \quad (1)$$

[$z^\lambda(\xi)$ is the coordinate of the path of integration, which is parametrized by allowing ξ to run from $-\infty$ to 0).]

* Work supported by the National Science Foundation.
[1] W. Ehrenberg and R. E. Siday, Proc. Phys. Soc. (London) **B62**, 8 (1949).
[2] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
[3] Y. Aharonov and D. Bohm, Phys. Rev. **123**, 1511 (1961).
[4] B. De Witt, Phys. Rev. **125**, 2189 (1962).
[5] Y. Aharonov and D. Bohm, Phys. Rev. **125**, 2192 (1962).
[6] F. J. Belinfante, Phys. Rev. **128**, 2382 (1962).

Belinfante,[6] however, regarded the integration over an arbitrary path as unsatisfactory, and instead suggested another gauge transformation, leading to a similar theory, in which the usual vector potential is replaced by

$$A_\mu''(x^\alpha)=\int F_{\eta\mu}(z^\lambda)\frac{(x^\eta-z^\eta)}{4\pi r^3}d^3z. \qquad (2)$$

Here the integration is carried out over a certain three-dimensional space-like hypersurface in the four-dimensional space-time continuum. The advantage of this transformation is that no arbitrary path of integration needs to be chosen. Rather, the three-dimensional space-like hypersurface can be taken to correspond to the coordinate frame of some observer. Of course, this means that the theory is not *manifestly* covariant, though (as we shall go into later) it may perhaps still be covariant in its over-all content.

Both De Witt and Belinfante then suggest that one can start with the field operators, $F_{\mu\nu}(z^\lambda)$, as a basic set of "observables," assuming the usual "local" commutation relations, between field quantities. [That is, field quantities at different places and at the same time all commute.] These will guarantee the *localizability* of the $F_{\mu\nu}(z^\lambda)$ "observables." On the basis of this result, they argue that one has obtained a suitable formulation of quantum electrodynamics, embodying the feature of localizability without using the potentials. As both De Witt and Belinfante recognize, however, the field equations now involve the *nonlocal* operators, $A_\mu'(x^\alpha)$ or $A_\mu''(x^\alpha)$, as given in Eqs. (1) and (2).

In our answer to De Witt, we pointed out that the transformations described above, are, in fact, obscuring the problem of what is meant by localizability. For to express this property naturally and in a clear fashion one requires not only local basic operators, but also local "interactions," not involving nonlocal functions of the basic operators such as those given in (1) and (2). In this way one obtains a clear and natural mathematical expression for the common "intuitive" notion that the world can be analyzed into "local" entities (i.e., the fields and charges) interacting by a sort of "contact" (which notion is, in fact, implicit in modern field theories).

Of course, one might argue that perhaps the above-described "intuitive" notion of localizability is no longer adequate, and that it may be desirable now to change this concept, in such a way that it will refer only to local field operators, without committing ourselves necessarily to local field equations as well. But (as we also noted in our answer to De Witt) there are serious, as yet unsolved, problems in trying to do this, because, *in general*, purely local commutation relations are not compatible with nonlocal field equations. (In other words, the propagation of initially "local" commutation relations by nonlocal equations of motion would *generally* lead to the development of nonlocal features in the commutation relations in question.) Therefore, the apparently "nonlocal" electrodynamical field equations given by De Witt and Belinfante, are known to be capable of expression without self-contradiction, only when they are derivable by some transformation from a local set of equations that involve the potentials. It is thus seen that the peculiar and restricted types of forms, (1) or (2), in which the field quantities must enter the equations of motion, are in some rather obscure and indirect way implying the property of localizability. On the other hand, this property is clearly and directly exhibited, when one does not eliminate the potentials. The problem of the form of the equations is, therefore, not *merely* a question of convenience or taste, at least if one wants not only to calculate some result, but also to see clearly what are the usually *unasserted* physical-geometrical assumptions that go into determining the mathematical form of the theory.

To bring out this point more sharply, we can refer to an analogous problem in which physical motions are linearly superposable (e.g., a harmonic oscillator). By the substitution, $u=v^2$, we could obtain a *formally* nonlinear equation, but at the expense of making the meaning of the property of linear superposition rather difficult to see. (If it turned out that the nonlinear equations were suspected on reasonable grounds to be *generally* self-contradictory, unless reducible to linear ones by some substitution, the analogy would be even better.)

The physical principle of linear superposability of motions thus helps to determine the (linear) *mathematical form* of the equations of motion and dictates for this purpose the choice of v as a "natural" dependent variable. In a similar way, the physical-geometrical principle of localizability helps to determine the (localizable) mathematical form of the electrodynamical equations and dictates for this purpose the potentials as a "natural" electrodynamical set of variables (leading to local "observables" *and* local equations of motion).

In a further paragraph added in the proofs, De Witt objected in yet another way to the conclusion described above, saying that the mathematical form of current quantum-electrodynamical theories is "determined by experiment" (stating further on that the same is true of Maxwell's equations). We were, in fact, somewhat surprised to encounter such an objection. For it seemed as if De Witt felt that theoretical physicists simply considered the results of experiments, and then wrote down essentially uniquely determined mathematical equations (except perhaps for various possible transformations leaving their physical implications unchanged), and that this was all done without the need for introducing any principles, ideas, or assumptions, going at all beyond what can be seen directly in the experimental results themselves. If this is what De Witt intended to say, then it is evidently false.

Thus, even Newton, who at one point claimed not to make hypotheses, actually brought in by implications a great many assumptions about space, time, causality, the inertial frame, the existence of forces acting at a distance, etc., which did, in fact, play a crucial part in determining the mathematical form of his equations of motion, and which were *never* "determined by experiment." When some of these assumptions were later questioned, (e.g., in the theory of relativity and the quantum theory), room was made for very fundamental changes in the mathematical forms of physical theories. In a similar way, neither Maxwell's equations nor those of quantum electrodynamics are determined simply *from experiment alone.* Thus, when Maxwell originally proposed his equations, he introduced the "displacement current" as a *hypothesis*, which (among other things) led to a simple and coherent *mathematical form* for his theory. Later, this hypothesis was confirmed by experiment, but even now, the assumption, for example, of linearity is far from verified in full detail by experiment. As for quantum electrodynamics, its mathematical form was originally suggested largely as a natural application of quantum theory to field theory. Since then, it has received a few experimental confirmations (e.g., the Lamb shift), but these can hardly be said to determine *uniquely* the mathematical form of the currently accepted theories. Among the principles that are helping to effect such a determination one is, as we have pointed out, the localizability of electromagnetic interactions.

3. ON THE PROBLEM OF A COMPLETE COMMUTING SET OF LOCAL OPERATORS

The problem of localizability was the main point that we wished to discuss in our earlier articles on the potentials. However, we also went into the question of a complete set of commuting "observables" in field theory, and into that of whether theories involving field quantities alone could solve the problem of obtaining such a set in a satisfactory way. Here, we unfortunately made use of some common but loosely applied terms, which may perhaps have led Belinfante to attribute to us intentions that we did not have. The first of these loose terms arose in our reference to an acceptable "covariant" form of quantum electrodynamics. By implication, we took this to mean "manifestly covariant." In doing this, we, of course, never intended to suggest that current manifestly covariant treatments are free of objections, doubtful assumptions, and possible sources of self-contradiction. Rather, all that was meant was that, as yet, no other treatments had been shown to achieve even the limited degree of coherence in the manner of leaving out infinities that has been obtained in manifestly covariant theories. Belinfante,[6] however, has argued that in an earlier article[7] he did actually prove the relativistic invariance of a non-manifestly-covariant theory, essentially equivalent to that resulting from the formulation described in connection with Eq. (2), of the present paper. Nevertheless, in a somewhat later paper,[8] he recognizes that the problem of getting rid of the infinities in a covariant way is yet to be dealt with. In his recent paper,[6] Belinfante suggests that this is due only to the tediousness of the requisite calculations, and to the fact that no one has cared to work out consequently what would happen if the calculations were actually done. This is, however, a matter of opinion. When we wrote our paper,[5] it seemed to us that there were serious difficulties in the way of such a program, and as yet, these difficulties do not seem to have been resolved. Therefore, our implicit assumption that all satisfactory covariant theories are also manifestly covariant was not basically wrong; although, admittedly, it would have been clearer if we had explicitly stated that we meant "manifestly covariant."

At this point, it will perhaps be well to make it clear that we did not (and do not) regard theories involving field quantities alone as of no potential interest or value. First of all, of course, it is necessary to see whether such theories can really deal with infinities in a satisfactory covariant way. But even if it should turn out that they can, it still seems to us that they must also face the problem, described earlier, of expressing the principle of localizability in a clear and natural way. Until they do this, it seems to us that they are obscuring an important physical-geometrical question without giving any corresponding gains to compensate for this loss. Naturally, it is quite possible that further research on these lines will resolve such problems, and lead to new results of interest. But for the present, it seems that potentials are still needed to play the "essential" part to which we have been referring.

4. ON THE PROBLEM OF SUBSIDIARY CONDITIONS IN ELECTRODYNAMICS

Now De Witt and Belinfante do suggest that there is *some* advantage to be gained by starting with the field operators, rather than with potentials, i.e., one can eliminate the subsidiary conditions (such as $\partial A_\mu/\partial x^\mu = 0$ in the Lorentz gauge). They seem to feel that the subsidiary conditions are creating serious and basic difficulties in quantum electrodynamics, so that it is an important advantage to be able to eliminate them. However, for reasons that we shall give, we do not think that the subsidiary conditions are really creating such serious problems, so that we do not regard the *mere* elimination of these potentials as a very significant gain. On the contrary, it seems to us that, as we shall try to show, the potentials may be furnishing important clues as to further developments in physics, so that their elimination may lead to obscure these clues (whatever may be the, as yet, unclear advantages, which this elimination might later be shown to have).

The main current objection to the use of subsidiary conditions is probably that one begins with a set of operators whose eigenfunctions include what are called "nonphysical" states, that do not satisfy the subsidiary condition. If it were not for the divergences and generally infinite results implicit in current theories, this

[7] F. J. Belinfante, Phys. Rev. **84**, 541 (1951).
[8] F. J. Belinfante, Phys. Rev. **91**, 1285 (1953).

would, of course, cause no difficulty of a purely mathematical nature. For one could simply consider the total set of solutions for the electrodynamical system (physical and nonphysical) and then choose those satisfying the subsidiary condition. But because current theories are divergent, special care is needed to avoid the inclusion of transitions to the "nonphysical" states in the computation of quantities, such as the Lamb shift. Such precautions require the addition of further injunctions, beyond what follows necessarily from the original theory. Belinfante[6] regards such devices as a kind of "trick," or "artifice," which "smuggles" in extraneous factors. Of course, Belinfante is quite correct in this. But then, is not the whole renormalization technique likewise just such an "artifice," which is being used as a kind of very useful and necessary but (we hope) temporary "stop-gap," until a new nondivergent theory can be developed?

Indeed, there is good reason to suppose that the current "stop-gap" theory is not even logically coherent, in the sense that when carried beyond a certain very limited range of operations and calculations, it may lead to self-contradictory results.[9] And if a theory is self-contradictory, then from problems that arise in dealing with any of its parts (such as the subsidiary conditions) we cannot draw any secure conclusions as to the real origin of the difficulties. (For example, by introducing the self-contradictory statement, $2=1$, somewhere in the theory, we might obtain absurd results in certain places, while in other places this contradiction would not affect the results. But none of the discussion would throw any reliable light on the validity of any particular part of the theory). Indeed, all that can be said thus far is that the problem of the subsidiary condition is not known as yet to introduce any mathematical difficulties beyond those that may be removed in some future theory that avoids the admittedly self-contradictory infinities of current theories. (In the Appendix, we suggest some more definite reasons why the elimination of these infinities may help us to avoid problems connected with subsidiary conditions.)

5. ON THE "OBSERVABILITY" OF THE POTENTIALS

Now, in the papers of De Witt and Belinfante, there seems to be still a further kind of objection to potentials, viz., that they are not "observable." And here, there is a type of confusion of terminology, which has in effect been built into the quantum theory itself. For the phrase "Hermitian operator whose eigenfunctions form a complete basis for the wave function" has been identified with the word "observable." So widespread is this use of the term, "observable," that it has become almost an unconscious habit among physicists. In fact, we ourselves used it in our reply[5] to De Witt. We wish here to apologize for furthering this confusion of terminology, and to make amends by trying to clarify the question a bit.

Now the common usage of the term "observable" is equivalent to a kind of unstated assumption: In every theory, there will be a complete set of wave functions, which are eigenfunctions of a complete commuting set of Hermitian operators, and *all of these operators are "observable."*

Where does this assumption come from? Of course, the assumption of linear operators grew up in the original largely mathematical theories of Heisenberg and Schrödinger, laying the basis of the quantum theory. But what is the origin of the further assumption that all operators of the type described above must be observable?

The full answer to the above question is perhaps as yet unclear. Nevertheless, one can see a number of significant reasons leading physicists to stress the importance of "observables." Historically, there was the famous principle of "Occam's razor" which enjoins upon us the desirability of not multiplying basic entities and concepts in an arbitrary way, particularly when they cannot be verified through observations. Then in the development of relativity, the criticism of the unobservable absolute space and the ether was an important means of freeing thought to allow the consideration of new ideas. Partly as a result of this development, there was a growing feeling that one should avoid the discussion of "unobservables" in physical theories. And when the quantum theory came into existence, this feeling may well have been one of the main factors leading to the identification of "basic Hermitian operators" with "observables."

While the principle of trying to avoid "unobservables" was in many ways helpful in the recent growth of physics, there has arisen what appears to be a harmful tendency to regard this principle as an absolute truth. To do this is to be caught in a contradiction; since there is no "observable" way of showing that this *is* in fact an absolute truth. Perhaps it is only a principle that should at the very least be used with "common sense" within certain limits. However, it may be possible to go even further, and to suggest that "unobservables" may never be *totally* eliminated in our thinking about physics. Indeed, in spite of the strong effort to get rid of "unobservables" in modern physics, there has been no dearth of them in recent physical theories. However, they are no longer called "entities." But instead, physicists now call them "mathematical quantities" or "conceptual aides." Thus, in quantum theory, everything has depended on the wave function, which is not, on the whole, a very observable sort of thing. Then there are the purely conceptual "waves" and "particles" in Bohr's principle of complementarity, which are also not "observable" though they are important in making clear the meaning of the theory. Then there are the "virtual states," "resonance," and a tremendous number of unobservable concepts that physicists find it

[9] L. von Hove, Physica **18**, 145 (1952); A. S. Wightman, Phys. Rev. **98**, 812 (1955).

convenient to talk about when they want to understand the results of the quantum theory and communicate with other physicists. Besides, in relativity, we have the unobservable coordinate frame, etc. As long as we call it a "mathematical quantity," or a "conceptual aide," physicists do not generally seem to have any real objection to working with unobservables. Indeed, as we have suggested above, it seems that *whatever the reason may be*, it is almost impossible to do theoretical physics without at least thinking of something unobservable, which makes the whole set of results cohere, hang together, and fit in some rationally comprehensible way.

So if we are making a theory of electrodynamics, can there be any real objection to calling the vector potential a "mathematical quantity" and a "conceptual aide" which helps us *clearly* to formulate the problem of localizability? The "nonobservable" aspects of these operators cause no real trouble, since the assumption that all such operators must be "observable" in the usual sense is, in any case, arbitrary. Just as the wave function has some not fully observable features, such as its phase, so we introduce Hermitian operators for the vector potential, that are in some similar sense not fully observable. After all, what can be asked of a theory except that it be logically coherent, fruitful, easily comprehensible as to what are the basic general principles entering into the determination of its form, and in agreement with experiments that have been done with the object of testing it? To ask that we reject *all* "unobservable" Hermitian operators seems a rather arbitrary further requirement. Although this requirement is commonly adopted in the entirely laudable effort to avoid philosophical preconceptions the fact is that in effect, it merely amounts to another philosophical notion which may be called the "antiphilosophical philosophy"; and this philosophy is even more arbitrary than the ideas it seeks to avoid.

But the introduction of "unobservables" often turns out to be more than a mere convenience in thinking. Sometimes (though not always) it serves as a clue to genuinely new developments. This is not surprising, since what is new begins by being *unknown* and, therefore, must escape the net of the kinds of things that are "observable." To suppose that all that we discuss must be "observable" is indeed equivalent to assuming that we ought already be able to recognize the general forms of everything that can be found in nature, and that only the "details" remain to be filled in. But if there is something really new, it may not "fit" into what we can now observe and recognize at all. Very often, however, we can get a vital clue to the unknown by seeing what is needed to give us a coherent and clear *conception* of certain phenomena that are already known. And here, it is very likely that something *unobservable* (at least for the present) will have to be considered.

There are many examples, but the authors would like to consider the Hamilton-Jacobi theory of classical mechanics. Mathematically this introduced an action function, S, which suggested, physically, a wave front, normal to the particle trajectory. The consideration of a whole wave front suggested an ensemble (perhaps statistical) of particle trajectories. Now, in classical mechanics, this wave front was considered to be totally "unobservable," since it was supposed that only the particle really existed. But the fact that the Hamilton-Jacobi theory gave such a *natural* expression for the classical canonical transformation might, *in principle*, have suggested (though it actually did not) that there was really some kind of a "wave" associated to the particles. If this were the case, the action function, S, would be proportional to the "phase" of this wave. This function S (or at least its *changes*) would, *in principle*, be "observable," when one would be able to do new kinds of experiments (e.g., interference) on the wavelike aspects of the system. And indeed, with the advent of quantum theory, we are, in fact, "observing" differences of phase, which did not even have *any meaning* in terms of classical "particle" conceptions.

Now, we propose that as the S function lent further mathematical coherence to the canonical formation of classical physics and, in principle, gave a clue (not actually used) toward quantum phenomena, so the electromagnetic potentials give coherence to the laws of electrodynamics, and in principle yield clues to further new phenomena. But now we may ask, "Why repeat the mistakes of the past? Why not try to pay some attention to the *clues* provided by the potentials?" In this short article, we cannot go into these clues in much detail. However, what is suggested is that the potentials are very important in expressing some of the *topological* and *structural* relationships that are behind the concept of *localizability* of charges and fields. (Some preliminary suggestions concerning the further significance of potentials are given in more detail elsewhere,[10,11] as well as in the Appendix of the present article.) This is of possible physical interest, first because if one is going to give up localizability (e.g., to get rid of infinities) it is desirable to see with some clarity and depth of understanding what one is giving up. Secondly, since the potentials may be related to qualitatively new topological and structural properties, they may actually be "observable," but only with correspondingly new (and as yet unknown) kinds of experiments (as during the period of classical physics, the action function, S, was not "observable," but became "observable" when electron interference experiments were first done).

[10] D. Bohm, in *Proceedings of Eighth International Congress on Low Temperature Physics*, (Butterworths Scientific Publications Ltd., London, 1962). A discussion of some of the topological and structural ideas is given here. See also, *The Scientist Speculates*, edited by I. J. Good (William Heinemann, London, 1962), p. 302.

[11] Y. Aharonov has shown that, in a quantum state in which the potentials are not well defined, potential differences at a given point in space may have a physical significance, manifested in the possibility of defining the velocity and position of a charged particle at the same time. Thus, one does not need to restrict physically significant functions of the potentials to circuit integrals, that are reducible to integrals of fields inside the circuit. A manuscript on this subject is now being prepared.

Mathematically speaking, then, it seems that in the interests of clarity of thought concerning quantum theory, it would be advisable to give up the unstated assumption that all "basic" Hermitian operators must be observable. When one looks into the question, one sees that this assumption really plays no part in current theories. For all the results of these theories would follow without such an assumption, merely from the supposition that *some* Hermitian operators are observable. We can leave open the question as to which of them are observable and which are not. Moreover, operators that are not observable in terms of what can be known and recognized at present may be related to observation in new ways, based on further laws and concepts, which we might come to learn at some future date.

Thus far, the very form of quantum mechanics seems to have required the assumption of a complete commuting set of basic Hermitian operators. Among these, for operators that are "observable" (by currently known means) the usual rules can be assumed to give the probability of a certain result as the square of the absolute value of the corresponding normalized eigenstate in the expansion of the wave function. But for operators that are not "observable" in the above-described sense, there is no need to assume a probability function at all, e.g., as a "positive definite" metric in Hilbert spaces. One can assume nondefinite metrics,[12] or one can simply leave the question of whether or not there is a metric of any kind in Hilbert space completely open. As to what these assumptions mean in a "physical" sense, this question too can be left open. Perhaps they may be interpreted later in terms of concepts that simply would not be "recognizable" when expressed in terms of current physical and mathematical conceptions of the quantum theory (as would happen with the interference properties of electrons, if we tried to state them in the classical "language" of well-defined particle trajectories).

Finally, we do not even wish to insist that the features of the quantum theory described above will not eventually have to be changed fundamentally. But it does seem desirable that we try to see clearly what we are assuming, what we are changing, and why we are doing so. We have gone rather extensively into these questions, because it seems that it is important at this stage to try to make them as clear as possible; and because lack of clarity on these points seems to be one of the reasons why what we have said about potentials has apparently not been sufficiently understood.

APPENDIX. SOME REMARKS ON THE RELATIONSHIP BETWEEN GAUGE INVARIANCE AND LOCALIZABILITY

In this Appendix, we add some remarks on the relationship between gauge invariance and localizability of basic field quantities, in the sense discussed in Sec. 2.

[12] S. N. Gupta, Proc. Phys. Soc. (London) **A63**, 681 (1951).

It is well known that there is a close analogy between problems arising out of gauge invariance in electrodynamics and certain other problems that arise from the covariance of the gravitational field equations in general relativity (see, for example, the interesting and systematic study of this subject by Bergmann and his co-workers[13]). It will be instructive first to consider briefly the relationship of such invariance problems to the question of localizability in gravitational theory, and then to go on to consider similar questions connected with gauge invariance in electrodynamics.

General relativistic covariance implies that the field equations for the gravitational potentials, $g_{\mu\nu}$, satisfy four identities (Bianchi identities), so that, in effect, four of the potentials are not determined by the initial conditions and the equations of motion. It is this lack of determination which leaves room for an arbitrary coordinate transformation, within the framework of the gravitational equations. Similarly, in electrodynamics, the conservation of charge follows from the field equations as an identity, with the result that the theory leaves room for invariance of physical results to an arbitrary gauge transformation of the potentials, which is similar to (but simpler than) the transformation of gravitational potentials brought about by a change in the coordinate frame.

Now, as we have indicated in this paper, it seems likely from current difficulties in field theories that at very short distances, the laws of physics will have to change fundamentally, in such a way as to alter the notion that basic field quantities are localizable. In some rough sense, it could be said that there is probably a "fundamental length," below which the basic concepts of current theories cease to be relevant. We have, in fact, already discussed the suggestion[10] that in the domain of the such short distances, space-time will have to be described in some inherently structural and topological sense, because it may possess a kind of irreducible "graininess," which denies the notion of the perfect continuity, that is built into current field theories.

It has often been proposed that the "fundamental length" is of the order of 10^{-13} cm, but we are inclined to favor the assumption that it will be of the order of 10^{-33} cm, the basic "gravitational" unit of distance (because such a fundamental length will signalize failure of ordinary notions of metric, and should therefore be deeply related to *gravitational* effects, rather than to electrical or nuclear interactions, which are the main factors at 10^{-13} cm).

If there should be a basic unit of length as described above, then current theories can have only some limited degree of covariance to arbitrary coordinate transformations. Roughly speaking, theories will be covariant only under general transformations in which the scale of length is not changed so much that physical structures such as nuclei are shrunk down to the size of 10^{-33} cm

[13] P. G. Bergmann and A. Janis, Phys. Rev. **111**, 1191 (1958).

(as large-scale thermodynamic laws are not invariant, if objects are shrunk to atomic dimensions in a transformation). We are, therefore, led to suggest that the Bianchi identities which follow only for complete invariance, do not hold *absolutely*, but only in some very high degree of approximation. If this is the case, then the corresponding auxiliary conditions in the gravitational potentials will also only be a very good approximation. All ten gravitational potentials will then have *some* physical significance. The four that should be irrelevant in a continuous space will in some indirect sense, represent the fundamental unit, underlying the discrete structure of space-time. In this way the assumption of a basic "graininess" to space-time, a step that will probably eliminate the infinities of current field theories, is seen also to get rid of the auxiliary conditions in the gravitational potentials which create such complicated problems in the quantum theory. In place of auxiliary conditions determined by certain operators, there will be equations implying a quantum state, in which there is a probability distribution of these operators, which would be a delta function, if space-time were continuous. However, if, as we assume, it is not continuous, the width of this probability distribution will determine the relationship of the coordinate frame to the basic unit of length (i.e., it will imply the size of the basic "grains," as specified in terms of the given coordinate frame).

A similar point of view can be adopted with regard to electrodynamics. Thus, we need not regard the subsidiary conditions as satisfied exactly, but rather, only to a very high degree of approximation. Then we can, for example, avoid the problem that the wave function is a delta function of $\partial A_\mu/\partial x^\mu$. Now, it will be spread over a finite but small range of this operator. Of course, charge will not, in general, be conserved, if this is assumed. But the failure of conservation may be very small, such that transitions in which the charge "quantum number" is changed will be much weaker than those in which isospin and strangeness are not conserved, so weak in fact that the phenomenon of charge nonconservation has not yet been observed.

We can illustrate this idea briefly by first considering the ordinary Lagrangian for the electromagnetic field:

$$L_1=\frac{\mathcal{E}^2-\mathcal{H}^2}{8\pi}+\mathbf{j}\cdot\mathbf{A}-\rho\phi,$$

with

$$\mathcal{E}=\frac{-1}{c}\frac{\partial\mathbf{A}}{\partial t}-\nabla\phi,\quad \mathcal{H}=\nabla\times\mathbf{A},$$

and with units in which $c=1$.

We then add a very small term

$$L_2=\Omega^2/2\lambda,$$

where $\Omega=\mathrm{div}\mathbf{A}+\partial\phi/\partial t$ and λ is a suitably adjusted constant. The total Lagrangian is then

$$L=L_1+L_2.$$

All components of the potential now have canonically conjugate momenta. Thus,

$$\mathbf{\Pi}=\mathcal{E}/4\pi$$

is canonically conjugate to $\mathbf{A}$, while

$$\Omega=(1/\lambda)(\partial\phi/\partial t+\mathrm{div}\mathbf{A})$$

is canonically conjugate to ϕ. The Hamiltonian is then

$$H=\frac{\mathbf{\Pi}^2+\mathcal{H}^2}{8\pi}+\left(\rho-\frac{\mathrm{div}\mathcal{E}}{4\pi}\right)\phi-\mathbf{j}\cdot\mathbf{A}+\frac{\lambda}{2}\Omega^2-\Omega\mathrm{div}\mathbf{A},$$

where ρ is the charge density and $\mathbf{j}$ is the current density. One readily obtains among the equations of motion, the following:

$$\nabla\times\mathcal{H}-\partial\mathcal{E}/\partial t=4\pi\mathbf{j}-\lambda\nabla\Omega,$$
$$\nabla\cdot\mathcal{E}=4\pi\rho+\lambda\partial\Omega/\partial t,$$
$$\Box\Omega=(4\pi/\lambda)(\partial\rho/\partial t+\mathrm{div}\mathbf{j}).$$

There is no subsidiary condition is this theory. But as $\lambda\to 0$, we readily verify that the wave function corresponding to the lowest state of excitation of Ω approaches a delta function of $\mathrm{div}\mathbf{A}+\partial\phi/\partial t$. In general, $\mathrm{div}\mathbf{j}+\partial\rho/\partial t\neq 0$ so that charge and current are not conserved, but as $\lambda\to 0$, $\mathrm{div}\mathbf{j}+\partial\rho/\partial t\to 0$ and the equations for $\mathcal{E}$ and $\mathcal{H}$ approach Maxwell's equations. The width of the probability distribution in Ω will be small enough so that $L_2=\Omega^2/2\lambda$ will approach zero as $\lambda\to 0$.

We note that the equations for $\mathcal{E}$ and $\mathcal{H}$ have, as "effective sources" for the fields, the quantities $4\pi\mathbf{j}-\nabla\Omega$ and $4\pi\rho+\partial\Omega/\partial t$. So in this theory, the electromagnetic field is distributed as if it had some other source than the particle positions (which latter do not in general give rise even to a perfectly conserved charge current). In this way, the "self-field" of a particle is effectively "generated" at some distance from the particle, so that its reaction back on the particle need not be as large as it is in the usual theory.

We see, then, that giving up gauge invariance is equivalent, indirectly, to altering some of the assumptions on localizability of interaction of charge and current. However, to take full advantage of the possibility of doing this, it is necessary also to give up perfect conservation of charge, as would, in any case, be only natural, if we regard this theory as the precursor of a further development, in which the "graininess" of space-time would be taken into account. For in a discrete space-time a differential equation for conservation of charge would, in any case, have no meaning.

It seems to be possible to obtain a further insight into the relation of the possible "graininess" of space-time to the factor of gauge invariance of electrodynamics by considering some proposals of Yang and Mills[14] and Utiyama.[15] These authors introduce an abstract isospace such that electrodynamical gauge invariance is a consequence of invariance of the theory to certain transformations in this space. In particular, a gauge transformation corresponds to the "independent" rotation of neighboring regions of isospace, so that gauge invariance expresses the complete "localizability" of the influence of fields in indefinitely small regions of isospace. The analogy to general relativity is that covariance to space-time transformations implies a similar possibility of "independent" deformation of each small region of space-time in such a transformation, thus expressing the indefinite "localizability" of the effects of fields in real space-time. (If the isospace is extended to several dimensions, then one can also subject strangeness and isospin properties to a similar treatment, thus opening the way to a more general approach to the theory of the quantum numbers of elementary particles.)

Thus far, we have considered only an *analogy* between isospace and the space-time of general relativity. But with the aid of some of the notions suggested in a more recent work,[16] it is possible to go further and to bring in a *definite relationship* between these two spaces. For the abstract isospace has been interpreted in terms of a real space-time structure of elementary particles. (In fact, it is just this concept of structure that has been developed further in the topological theories[10] to which we referred earlier in the present paper.)

It follows then that if there is a basic "graininess" to space-time, this may have implications not only for general relativistic transformations, but also for isospace transformations, which reflect the behavior of structures in relativistic space-time. Roughly speaking, over a region of the size of a "grain" one can no longer allow "independent" rotations of smaller subregions, because the notion of such a subregion no longer has its usual meaning. Thus, there is established a limit both to general relativistic covariance, and to invariance to transformations in isospace and, therefore, to gauge invariance and detailed conservation of charge. As a result it becomes possible to give a physical significance to all four components, A_μ, of the potential, which now determine not only the field quantities $\mathcal{E}$ and $\mathcal{H}$ in the usual way, but also, in a more indirect way, the "scale factor," below which continuous notions of space-time and detailed conservation of charge cease to apply. At the same time, current difficulties associated to the subsidiary conditions are evidently bypassed as irrelevant, since these conditions are now replaced by well-defined relationships satisfied by the quantum state of the whole field.

It is evident that these proposals constitute a particular case of the suggestion in Sec. 5 of this paper to regard the potentials as *clues* to some new features of space-time and properties of charge. In this way, one may perhaps obtain insights into the reasons why potentials seem to be "essential" for expressing the property of *localizability* in a simple and natural way.

A more systematic development of such a point of view is now being carried out, in which the "graininess" of space-time is interpreted in terms of topological and structural concepts. It is hoped that this will soon be ready for publication.

[14] N. Yang and R. L. Mills, Phys. Rev. **96**, 1911 (1954).
[15] R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[16] L. de Broglie, D. Bohm, F. Halwachs, P. Hillion, T. Takaboyasi, and J. P. Vigier, Phys. Rev. **129**, 438 (1963).