GAUGE TRANSFORMATIONS AND QUANTUM MECHANICS (III):
THE CONVENTIONAL INTERPRETATION OF QUANTUM MECHANICS

Kuo-Ho Yang
Department of Physics and Astronomy
University of Maryland, College Park, Maryland 20742 U.S.A.

August 1976

U. of Md. TR #77-019
U. of Md. PP #77-038

MASTER

UNIVERSITY OF MARYLAND
DEPARTMENT OF PHYSICS AND ASTRONOMY
COLLEGE PARK, MARYLAND

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

GAUGE TRANSFORMATIONS AND QUANTUM MECHANICS (IIT):
THE CONVENTIONAL INTERPRETATION OF QUANTUM MECHANICS

Kuo-Ho Yang*
Department of Physics and Astronomy
University of Maryland, College Park, Maryland 20742 U.S.A.

August 1976

ABSTRACT

The gauge problem of the conventional interpretation of quantum mechanics is examined carefully within the framework of treating electromagnetic fields as classical and external quantities. It is found that the conventional interpretation not only is gauge dependent but also does not allow the degree of freedom of choosing a gauge in the Schrödinger equation.

The fundamental approach to the gauge problem adopted in this paper is Lamb's criterion which states that the "interaction" Hamiltonian in a gauge can represent the true physical interaction if and only if the theoretical probabilities constructed "in the usual way" in the same gauge are the physically true probabilities. Imposition of necessary conditions of an observable on the theoretically constructed probabilities leads to constraints on the gauge in the Schrödinger equation. In particular, consideration of causality excludes the complete α -Lorentz gauge (defined by $\nabla \cdot \mathbf{A} + \frac{1}{\alpha^2 c} \frac{\partial \phi}{\partial t} = 0$) with $\alpha \neq 1$ and consideration of the conservation laws excludes the Lorentz gauge.

*Supported in part by the National Science Foundation.

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Energy Research and Development Administration, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

I. INTRODUCTION

This paper is the third paper in the series of our investigation on gauge transformations and quantum mechanics. In our first paper [1], we have formulated a gauge invariant interpretation of quantum mechanics by interpreting operators according to their equations of motion. Our second paper [2] deals with the propagation characteristics of potentials in a restrictive class of gauges. There are two objectives in that paper. The first objective, by far the most important one, is to study the feasibility of gauge invariance of and possible problems in the quantization of only the electromagnetic radiation fields by quantizing the potentials. The second objective is to pave the way for the main theme of this third paper: to study the gauge dependence of the conventional interpretation of quantum mechanics.

In our first paper, the most important point as far as probability amplitude are concerned is our recognition that once a gauge (more specifically, a set of potentials) is chosen to construct a Hamiltonian to be used in the Schrödinger equation, the basis functions by which probability amplitudes are defined cannot be arbitrary. Furthermore, the basis functions are the eigenfunctions of the particle's energy operator whose time derivative is defined to be the power operator due completely to external time-varying fields. In other words, the basis functions are a result of the gauge initially chosen and the degree of freedom of choosing the basis functions in a way independent of our initial gauge therefore does not exist.

The above consideration is, of course, somewhat "strange" when compared with our traditional understanding of quantum mechanics. In our traditional understanding, not only can we choose the basis functions to be the eigenfunctions of the "unperturbed" Hamiltonian but also we have

3

the degree of freedom of choosing a gauge in the Schrödinger equation. That is, our traditional understanding tells us that we have two degrees of freedom in the interpretation of quantum mechanics: the first freedom to choose the basis functions and the second freedom to choose a gauge.

This two-degree-of-freedom thesis was first challenged by Lamb [3]--substantiated with experimental evidence. From the resonance curves of the quenching of the hydrogen atomic metastable state by RF fields, he observed that "the usual interpretation of probability amplitudes"¹ was gauge-dependent and was correct only in the gauge in which the "interaction" Hamiltonian was of the dipole form $\vec{r} \cdot \vec{E}$, rather than the "usual" $\vec{A} \cdot \vec{p}$ (under the long wavelength and lowest order approximation). This observation has become the subject of several theoretical investigations [4-6] that are completely within the framework of the conventional formulation. The net result of these investigations is that we have different and incompatible versions of explanation.²

How to interpret Lamb's "simple" message is, we believe, the most crucial point of the whole thing. We wish to state our opinions below. First, Lamb's results concerning $\vec{E} \cdot \vec{r}$ vs. $\vec{A} \cdot \vec{p}$ indicate to us that we must re-examine very carefully the usual arguments of either "proving" that "the usual interpretation of probability amplitudes" is "gauge invariant" or "proving" that a certain "special" gauge is "desired." Furthermore, the criterion or criteria used to arrive at these conclusions must not be perfect. Secondly, we believe that if Lamb's words are taken seriously, we must also investigate the higher multipole interactions.³ This is because being able to explain why $\vec{E} \cdot \vec{r}$ is "more correct" than $\vec{A} \cdot \vec{p}$ completely within the framework of the conventional formulation can not and does not tell us what to do with higher multipole interactions

4

which are mainly the result of expanding $\vec{A} \cdot \vec{p}$, not $\vec{E} \cdot \vec{r}$. Consideration of the above leads to our last interpretation of Lamb's observation. That is, without investigating the gauge problem in a mathematically rigorous way under the circumstance where no approximation is allowed, we believe Lamb's result cannot be explained in a flawless and perfect way. It is in the spirit of this last opinion that this paper differs in all respects from the references quoted in Refs. [4] to [6].

There are two important objectives that we wish to accomplish in this paper. Both are related to Lamb's conclusion on $\vec{E} \cdot \vec{r}$ vs. $\vec{A} \cdot \vec{p}$ although neither is directly involved. (For a derivation of the electric dipole interaction, see Ref. [1]). Our first objective is to understand the exact reasons why "the usual interpretation" is not gauge invariant. This, of course, will certainly involve clarifying some of the confusion deeply embedded in the usual arguments that "prove" that "the usual interpretation" is gauge invariant.

Our second objective is to investigate what the best choice of a gauge should be. Here, we shall be closely guided by the concept used by Lamb in arriving at his conclusion on $\vec{E} \cdot \vec{r}$ vs. $\vec{A} \cdot \vec{p}$. Our emphasis here is to show explicitly that causality and the conservation laws can dictate the conditions on gauges in the Schrödinger equation when probability amplitudes are to be defined "in the usual way."

This paper is divided into five sections. In Section II, we re-examine the "usual" definition of probability amplitudes and its consequences. In Section III, we investigate the criterion or criteria on how to choose a gauge. Section IV is devoted to discussing the relation between "the usual interpretation" and causality and the conservation laws. Finally, in Section V, we summarize the important points of this paper and discuss our results.

II. THE CONVENTIONAL INTERPRETATION

A. Definition of "the Usual Interpretation"

Assume that we have a charged particle of mass m and charge e moving in the presence of a static, conservative force $e\mathbf{E}_0(\mathbf{r})$ and time-varying electromagnetic fields $\mathbf{E}(\mathbf{r},t)$ and $\mathbf{B}(\mathbf{r},t)$. Then the Hamiltonian H^g in gauge g is

$$H^g = \frac{1}{2m} \left(\mathbf{p} - \frac{e}{c} \mathbf{A}^g \right)^2 + eV_0(\mathbf{r}) + e\phi^g(\mathbf{r},t) \quad (II.1)$$

$$= c\alpha \cdot \left(\mathbf{p} - \frac{e}{c} \mathbf{A}^g(\mathbf{r},t) \right) + \beta mc^2 + eV_0(\mathbf{r}) + e\phi^g(\mathbf{r},t)$$

where

$$eV_0(\mathbf{r}) \equiv - \int_{\infty}^{\mathbf{r}} e\mathbf{E}_0(\mathbf{s}) \cdot d\mathbf{s}. \quad (II.2)$$

The potentials $\mathbf{A}^g$ and ϕ^g are potentials in gauge g and generate the time-varying fields $\mathbf{E}(\mathbf{r},t)$ and $\mathbf{B}(\mathbf{r},t)$ through the relations

$$\mathbf{E}(\mathbf{r},t) = - \nabla \phi^g(\mathbf{r},t) - \frac{1}{c} \frac{\partial}{\partial t} \mathbf{A}^g(\mathbf{r},t), \quad (II.3)$$

$$\mathbf{B}(\mathbf{r},t) = \nabla \times \mathbf{A}^g(\mathbf{r},t).$$

For convenience, we also assume that $\mathbf{E}$ and $\mathbf{B}$ never become static, except when they both vanish identically, during the time interval of concern.

The wave function in gauge g , ψ^g , satisfies the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \psi^g(\mathbf{r},t) = H^g(t) \psi^g(\mathbf{r},t) \quad (II.4)$$

with initial condition $\psi^g(\mathbf{r},t_1)$ which is also gauge dependent. In order to discuss the conventional interpretation more precisely, we define "the usual expansion coefficients" of ψ^g by^{4,5}

$$C_1^g(t) \equiv \langle \phi_1 | \psi^g(t) \rangle \equiv \int d\mathbf{r} \phi_1^*(\mathbf{r},t) \psi^g(\mathbf{r},t) \quad (II.5)$$

where $\{\phi_i(\mathbf{r})\}$ are the orthonormal eigenfunctions of the "unperturbed" Hamiltonian H_B^0 with corresponding eigenvalues $\{\epsilon_i\}$, i.e.,

$$H_B^0 \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}),$$

$$\langle \phi_i | \phi_j \rangle = \delta_{ij}, \quad (II.6)$$

where H_B^0 , in our situation, assumes the form

$$H_B^0 \equiv \frac{1}{2m} \mathbf{p}^2 + eV_0(\mathbf{r})$$

$$\equiv c\alpha \cdot \mathbf{p} + \beta mc^2 + eV_0(\mathbf{r}). \quad (II.7)$$

"The usual interpretation" in gauge g is represented by the definition

$$P_{1,\epsilon_1}^g(t) \equiv |C_1^g(t)|^2 \quad (II.8)$$

with the understanding that $P_{1,\epsilon_1}^g(t)$ is to be regarded as observable and is to be interpreted as the probability at time t of finding the charged particle in the state i that has energy ϵ_i .

The differential equations for $\{C_1^g(t)\}$ are

$$i\hbar \frac{d}{dt} C_1^g(t) = \epsilon_1 C_1^g(t) + \sum_j C_j^g(t) \langle \phi_1 | H_R^g(t) | \phi_j \rangle \quad (II.9)$$

where the "interaction" or residual Hamiltonian $H_R^g(t)$ is

$$H_R^g(t) \equiv H^g(t) - H_B^0. \quad (II.10)$$

Since the expansion coefficients $\{C_1^g(t)\}$ are to be interpreted as probability amplitudes, which means the probabilities $\{P_{1,\epsilon_1}^g(t)\}$ are to be constructed according to Eq. (II.8), "the usual matrix elements"

$\langle \phi_1 | H_R^g(t) | \phi_j \rangle_{i,j}$ are therefore to represent physical transitions in this interpretive scheme. Therefore within the framework of "the usual interpretation", all discussions and investigations center around the "physics" of "the usual matrix elements" of the "interaction" Hamiltonian $H_R^g(t)$.

difficulty

ies the way discussed above
e probability amplitudes in a
her set of potentials (A', ϕ')

$$\text{or } g \xrightarrow{\chi} g' \quad (\text{II.11})$$

(II.11) means

$$\chi(r, t) \quad (\text{II.12})$$

$$\chi(r, t),$$

nction of r and t ,

$$t), \quad (\text{II.13})$$

$$(4) \quad \chi(r, t) = \chi(r, t).$$

For future convenience, any function satisfying the conditions in Eq. (II.13) is called a gauge. Also for the sake of clarity, whenever we say that g' is different from g or symbolically $g' \neq g$, we mean exactly⁶

$$A^{g'}(r, t) \neq A^g(r, t) \quad (\text{II.14})$$

(Note that this usage of the term gauge is different from the one in classical electrodynamics. In classical electrodynamics, two sets of potentials are said to be in the same gauge if they satisfy the same gauge condition.⁷ This is not so when discussing gauge transformations in "the usual interpretation of probability amplitudes." This, we hope, should always be borne in mind.)

Now, the wave function $\psi^{g'}$ in gauge g' satisfies

$$i\hbar \frac{\partial}{\partial t} \psi^{g'} = H^{g'} \psi^{g'} \quad (\text{II.15})$$

where

$$H^{g'} = \frac{1}{2m} \left(p - \frac{e}{c} A^{g'} \right)^2 + eV_0(r) + e\phi^{g'} \\ = c\alpha \cdot \left(p - \frac{e}{c} A^{g'} \right) + eV_0(r) + \beta mc^2 + e\phi^{g'}. \quad (\text{II.16})$$

As can be easily shown, $\psi^{g'}$ and ψ^g are related by

$$\psi^{g'} = R(\chi) \psi^g \quad (\text{II.17})$$

where

$$R(\chi) \equiv \exp\left[\frac{ie}{\hbar c} \chi(r, t)\right]. \quad (\text{II.18})$$

Now, let us proceed to discuss how to construct the probability amplitudes and consequently the probabilities in gauge g' (i.e., from $\psi^{g'}$). By this we shall show that it is a nontrivial problem. Intuitively, we may try

$$c_1^{g'}(t) \equiv \langle \phi_1 | \psi^{g'} \rangle, \\ p_{1, \epsilon_1}^{g'}(t) \equiv |c_1^{g'}(t)|^2. \quad (\text{II.19})$$

The justification of this interpretation is as follows. If we regard gauge g , the original gauge where the basis functions are chosen to be $\{\phi_i(r)\}$, as an arbitrary gauge, then gauge g' should also be treated on the identical footing as gauge g . This is because

$$|\psi^g(r, t)|^2 = |\psi^{g'}(r, t)|^2 \text{ for all } r \text{ and } t, \quad (\text{II.20})$$

which shows that as far as the probability $P(r, t)$ as a function of r and t is concerned, there is no difference between these two gauges. Therefore

what is done in gauge g must also be done in gauge g' . This means that the basis functions in gauge g' must also be chosen to be $\{\phi_i(r)\}$. Otherwise, it is impossible to justify the original definition of probability amplitudes in gauge g by Eqs. (II.5) and (II.8) since gauge g is regarded as an arbitrary gauge here.

The trouble with the interpretation in Eq. (II.19) is that, in general, $\{p_{1,\epsilon_1}^{g'}(t)\}$ are not always identical to $\{p_{1,\epsilon_1}^g(t)\}$. Note that we require the probabilities, not probability amplitudes, to be identical. To see this, we first assume that

$$p_{1,\epsilon_1}^{g'}(t) = p_{1,\epsilon_1}^g(t).$$

This assumption, however, leads to

$$\langle \psi | R^+(x) H_B^0 R(x) - H_B^0 | \psi \rangle = 0,$$

which is not possible for any arbitrary gauge function x .

What we have just discussed indicates to us that if "the usual interpretation" in one gauge agrees with physically observed phenomena, then "the usual interpretation" in another different gauge generally disagrees with physical measurements. This means that one can choose the basis functions to be $\{\phi_i(r)\}$ only for certain gauge or gauges, but not for all gauges.

C. A "Remedy"

As we have just shown, the interpretation Eq. (II.19) in gauge g' is in general not consistent with the interpretation Eqs. (II.5) and (II.8) in gauge g . The main reason leading to this inconsistency is our insistence

in regarding any gauge as being arbitrary and in choosing $\{\phi_i(r)\}$ as the basis functions for all arbitrary gauges. This inconsistency, however, can be completely eliminated if we regard gauge g as being more "special" than gauge g' and devise a different way to construct the probability amplitudes in gauge g' . Once we have decided not to treat gauges g and g' on the same footing, there is then no reason to demand that $\{\phi_i\}$ be chosen as the basis functions for both gauges.

Upon this consideration, we now wish to construct a set of basis functions $\{\phi_i^X\}$ such that $\{\phi_i^X\}$ will be as complete as $\{\phi_i\}$ and such that the interpretation resulting from $\{\phi_i^X\}$ in gauge g' is consistent with the interpretation Eqs. (II.5) and (II.8) in gauge g . That is, we require⁸

$$c_{1,\epsilon_1}^{g',-X}(t) \equiv \langle \phi_1^X | \psi^{g'} \rangle \quad (II.21)$$

$$p_{1,\epsilon_1}^{g',-X}(t) \equiv |c_{1,\epsilon_1}^{g',-X}(t)|^2,$$

and

$$p_{1,\epsilon_1}^{g',-X}(t) = p_{1,\epsilon_1}^g(t) \quad (II.22)$$

for all i and t . This set of basis functions can be constructed easily by

$$\phi_i^X \equiv e^{i\alpha(t)} R(x) \phi_i(r) \quad (II.23)$$

where $\alpha(t)$, a function of time only, satisfies that its first order derivative exists and is continuous.

The most difficult question confronting us is whether Eqs. (II.21) to (II.23) constitute a valid proof of the gauge invariance of the interpretation Eqs. (II.5) and (II.8) in gauge g . In order to answer this question, we must recall again the circumstances that led to defining the probability amplitudes in gauge g' by Eq. (II.21). As we have observed before, if we insist that gauge g' be as arbitrary as gauge g , then $\{\phi_i(r)\}$

must be chosen as the basis functions in gauge g' since $\{\phi_1\}$ are chosen as the basis functions in gauge g . The mere fact that we use $\{\phi_1^X\}$ as the basis functions in gauge g' can be justified only if gauge g is more "special" than gauge g' . Therefore Eq. (II.21) automatically guarantees that we have already lost a degree of freedom in our interpretation in gauge g in the sense that gauge g can not be arbitrary. Once gauge g is not arbitrary, Eq. (II.22) only implies that the interpretation Eq. (II.21) in gauge g' is consistent with the interpretation Eqs. (II.5) and (II.8) in gauge g . It by no means implies that the interpretation in gauge g is gauge invariant. In other words, Eq. (II.22) only tells us that if $\{p_{1,\varepsilon_1}^g(t)\}$ are physically correct so are $\{p_{1,\varepsilon_1}^{g,-X}(t)\}$ and that if $\{p_{1,\varepsilon_1}^g(t)\}$ are physically wrong so are $\{p_{1,\varepsilon_1}^{g,-X}(t)\}$. In no way does Eq. (II.22) ever have any slightest indication on the physical correctness or incorrectness of the interpretation in gauge g . This point must always be borne in mind and understood since it is the interpretation in gauge g , not in gauge g' , which we shall investigate rigorously.

III. HOW TO CHOOSE A GAUGE

In the previous section, we have learned that if consistent interpretations are desired for two different gauges g and g' ($\neq g$), identical procedures must not be used to construct probability amplitudes in these two gauges. This immediately presumes that gauge g must be more "special" than gauge g' (or vice versa). The central problem now is to investigate the underlying physics that will serve as our guidance for the choice of a "special" gauge in which the probability amplitudes are to be defined "in the usual way" (i.e., according to Eqs. (II.5) and (II.8)). Once this "special" gauge is found, the interpretation in a different gauge must follow the procedure defined by Eqs. (II.21) to (II.23). Here, we first examine our "traditional" arguments.⁹

A. The "Traditional" Reasoning

Traditionally, the reasons behind any particular choice of gauges can be best understood from the differential equations for the probability amplitudes. In gauge g , these equations are

$$i\hbar \frac{d}{dt} C_1^g(t) = \epsilon_1 C_1^g(t) + \sum_j C_j^g(t) \langle \phi_1 | H_R^g(t) | \phi_j \rangle. \quad (\text{III.1})$$

For the "clarification" of the underlying physics hidden in any "interaction" Hamiltonian, $H_R^g(t)$ is generally written as

$$H_R^g(t) = -\frac{1}{c} A_\mu^g J_\mu \quad (\text{III.2})$$

where J_μ is the particle's charge-current operator. In nonrelativistic quantum mechanics, $H_R^g(t)$ assumes the explicit form

$$H_R^g(t) = -\frac{e}{mc} (A^g \cdot p + p \cdot A^g) + \frac{e^2}{2mc^2} A^g \cdot A^g + e\phi^g, \quad (\text{III.3})$$

while in relativistic quantum mechanics, it is

$$H_R^g(t) = c\alpha \cdot \left(-\frac{e}{c} A^g\right) + e\phi^g. \quad (\text{III.4})$$

From Eq. (III.1) it is clear that "the usual matrix elements" $\{\langle \phi_i | H_R^G | \phi_j \rangle\}_{i,j}$ govern the time evolution of $\{C_i^G(t)\}$. Therefore, if $H_R^G(t)$ represents the true interactions of the quantum system with external electromagnetic fields, $\{C_i^G(t)\}$ will certainly generate the physically true probabilities. (It is also clear from Eq. (III.1) that if $\{C_i^G(t)\}$ generate the physically true probabilities, it would then follow that $\{\langle \phi_i | H_R^G | \phi_j \rangle\}_{i,j}$ represent the true transitions.) Invariably, the emphasis of our "traditional" arguments is on the "interaction" Hamiltonians. That is, it is hoped that the gauge in which "the usual interpretation" is physically true can be obtained by investigating the physical correctness of the "interaction" Hamiltonians.

How to determine the physical correctness of any "interaction" Hamiltonian is the most crucial issue. This is where we find our "traditional" arguments very unsatisfactory. (For example, one "traditional" argument in favor of the Coulomb gauge goes like this: the "interaction" Hamiltonian in the Coulomb gauge represents the true interaction since the vector potential represents the "transverse" fields and since the interaction between the current and the transverse fields is the true interaction.¹⁰) This is because our "traditional" arguments lead us to believe, upon moments of reflection, that how to choose a gauge can not be derived from our physical interpretation of "the usual probability amplitudes." In other words, our traditional reasoning tells us that the gauge constraint, upon which "the usual interpretation" solely relies for its validity, must exist outside the realm of the Schrödinger equation, "the usual interpretation" (Eqs. (II.5) and (II.8)), and Maxwell's equations (since Maxwell's equations are gauge invariant). This result, as we shall see next, is exactly to the contrary of Lamb's criterion.

B. Lamb's Criterion

In 1952, as a by-product of the famous Lamb's shift measurements [3], Lamb observed that out of the two "interaction" Hamiltonians $-e\mathbf{A}\cdot\mathbf{p}/mc$ and $-e\mathbf{r}\cdot\mathbf{E}$, only $-e\mathbf{r}\cdot\mathbf{E}$ yielded results that agreed with experimental curves (when probability amplitudes are constructed according to Eq. (II.5)). From this, he concluded that "the usual interpretation of probability amplitudes" was correct only in the gauge of the electric dipole approximation [3] (for bound systems). In order to understand how this "simple" statement carries with it a new concept on how to choose the "special" gauge discussed in Sect. III.A, we must analyze very, very carefully the procedure by which Lamb arrived at his conclusion 25 years ago.

First, Lamb had the physically true (i.e., observed) probability of the hydrogen atomic metastable state after it was "quenched" by an RF field. This probability was plotted as a function of the frequency of the RF field. He then calculated the theoretical probability, also as a function of the driving frequency, using the "interaction" Hamiltonians $-e\mathbf{A}\cdot\mathbf{p}/mc$ and $-e\mathbf{r}\cdot\mathbf{E}$, both under the long wavelength and lowest order approximation. Comparing these two sets of theoretically predicted probability with the physically true probability, Lamb observed that the prediction calculated using $-e\mathbf{r}\cdot\mathbf{E}$ agreed with the observed probability, while that from $-e\mathbf{A}\cdot\mathbf{p}/mc$ did not. From this, he made his conclusion that "the usual interpretation of probability amplitudes" was correct only in the gauge of $-e\mathbf{r}\cdot\mathbf{E}$ and was not correct in the gauge of $-e\mathbf{A}\cdot\mathbf{p}/mc$.

(As a by-product of the "quenching" of the metastable state, it should also be pointed out that the metastable state was "quenched" in a region where no clear-cut definition of the near-zone or of the intermediate-zone is possible. But from wavelength consideration, it is amply clear that the

"quenching region" was certainly not in the radiation zone of the RF field. This complicates matters considerably for those who wish to explain or to contradict theoretically Lamb's conclusion. This is because any particular form of potentials explicitly used must be justified first within the framework of classical electrodynamics. (See Ref. [1] for a derivation that is completely independent of the explicit forms of the potentials.))

The above-stated procedure can be summarized in the following:

Lamb's criterion: "The usual matrix elements"

$\{\langle \phi_i | H_R^g(t) | \phi_j \rangle\}_{i,j}$ can represent the physically true interaction if and only if
 $\{p_{i,\epsilon_1}^g(t)\} \equiv \{|c_i^g(t)|^2\}$ are the physically true probabilities.

First, let us compare this with the basic reasoning behind the "traditional" arguments. As stated before, the "traditional" arguments assert that only if $H_R^g(t)$ represents the "true" interaction can $\{C_i^g(t)\}$ generate the true probabilities through Eq. (II.9). But, how to judge if $H_R^g(t)$ represents the "true" interaction can not be determined from within quantum mechanics. This is, however, not true with Lamb's criterion. Lamb's criterion, as stated above, allows us to determine if $H_R^g(t)$ represents the "true interaction" by whether $\{C_i^g(t)\}$ generates the true probabilities.¹¹ Therefore, it demands, in effect, that the gauge constraint be determined from within quantum mechanics. That is, the Schrödinger equation and the postulate of probabilities Eq. (II.8) can completely determine the gauge constraint.

Allowing the Schrödinger equation and "the usual interpretation" to completely decide the gauge constraint immediately leads to one inevitable consequence. This consequence is: whether or not A^g represents whatever

kind of photons should not and can not be used as the only reason to justify a certain choice of gauge.¹² This is one of many important conclusions of this paper.

Here, we must also point out that Lamb's criterion is a very reasonable criterion within the framework of quantum mechanics (i.e., when electromagnetic fields are treated as classical and external quantities). In this criterion, the "interaction" Hamiltonians need not be treated as observables; whether or not they are observables is irrelevant. The question of the validity of Lamb's criterion when electromagnetic fields are to be quantized is not within the scope of this paper and is also beyond this author's ability. (However, if classical physics is to be of any guidance in our understanding of quantum mechanics, we believe that Lamb's criterion will still be valid. It is because according to the "standard" physical interpretation of the Poynting theorem [8], $-E \cdot J$ serves both as the interaction between matters and EM waves and also as the source (or sink) of electromagnetic radiation energies. Furthermore, upon the long wavelength and lowest order approximation, $E \cdot J$ reduces to the dipole interaction.¹³)

One possible annoyance that may arise from Lamb's criterion as formulated above is the question of uniqueness. More explicitly, we are asking the question whether there exist two different gauges g and $g' \neq g$ such that

$$p_{i,\epsilon_1}^g(t) = p_{i,\epsilon_1}^{g'}(t)$$

for all i and t . (For the definition of $g' \neq g$, see Sect. II.B.) If this situation occurs, then Lamb's criterion will not be able to distinguish between these two "interaction" Hamiltonians. This situation, however, will become bothersome only if the probabilities in these two gauges are the physically true probabilities. We have no reason to worry about it if $\{p_{i,\epsilon_1}^g(t)\}$ and

$\{p_{1,\epsilon_1}^g(t)\}$ are not the physically true probabilities since we just simply conclude that both gauges are "wrong". (By a gauge is "wrong" (or "correct") we mean the probabilities constructed in this gauge according to Eq.(II.8) are not (or are) the physically true probabilities.) What concerns us in this paper is to use Lamb's criterion to locate "wrong" gauges; the question of uniqueness, therefore, should not affect us at all. (Nonetheless, a casual consideration of a rotation on a state vector (wave function) in the Hilbert space indicates that only when $g'=g$ can then $p_{1,\epsilon_1}^g(t) = p_{1,\epsilon_1}^{g'}(t)$. We shall not prove this statement since the outcome has no bearing on the consequences of this paper.)

How to apply Lamb's criterion in order to have some meaningful and also useful results becomes our concern now. Basically, one can use this criterion to find a "correct" gauge. This is exactly what Lamb has done. The only difficulty in applying this criterion in this way is that we must have the physically true probabilities first. But, physically true probabilities can only be obtained from experiments. This is not our primary interest here.

What we are interested in is to apply Lamb's criterion without the physically true probabilities. Under this circumstance, it is immediately clear that we can not use Lamb's criterion to locate a "correct" gauge. Furthermore, it is also impossible to judge whether a gauge is "wrong". Therefore, in order to do something meaningful under this difficult circumstance, we use the following substitute for the physically true probabilities. We define that $\{p_{1,\epsilon_1}^g(t)\}$ can not be the physically true probabilities if any of the requirements from relativity (on the part of the propagation of electromagnetic physical information) and from the conservation laws is violated. With this substitute, we can now start our work on locating the "wrong" gauges.

IV. "WRONG" GAUGES

In this section, we first review the propagation properties of the complete α -Lorentz gauge, then formulate a mathematical theorem, and finally apply this theorem and Lamb's criterion to investigate the physical consequences of "the usual interpretation" in the α -Lorentz gauge for all $\alpha > 0$.

A. Review of the Complete α -Lorentz Gauge

Before Lamb's criterion with our substitute for the true probabilities can be applied, we must have mathematically correct and physically unambiguous solutions for potentials in different electromagnetic gauges. This is where the propagation characteristics of the potentials, vector and scalar, of the complete α -Lorentz gauge become indispensable. Furthermore, we have defined in Sect. II B the meaning of gauges in quantum mechanics in a slightly different fashion. We must re-define our gauges in quantum mechanics in terms of electromagnetic gauges. All these will be briefly dealt with here. For a more detailed investigation of gauge transformations, see Ref. [2].

Assume that we have a localized charge and current distribution represented by $\rho_{ex}(r,t)$ and $J_{ex}(r,t)$, then Maxwell's equations for potentials A and ϕ are

$$\begin{aligned} \nabla^2 \phi + \frac{1}{c} \frac{\partial}{\partial t} (\nabla \cdot A) &= -4\pi \rho_{ex} , \\ \left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) A &= -\frac{4\pi}{c} J_{ex} + \nabla \left(\nabla \cdot A + \frac{1}{c} \frac{\partial}{\partial t} \phi \right) . \end{aligned} \quad (IV.1)$$

We further assume that this charge-current distribution is turned on at t_0 .

What we mean by "turning on" is the following:

$$\rho_{ex}(r,t) = J_{ex}(r,t) = 0 \quad \text{for all } r \text{ and all } t \leq t_0 , \quad (IV.2)$$

and there exists at least an r where either $\rho_{ex}(r,t)$ or $J_{ex}(r,t)$ or both do

not vanish if, however small ϵ is,

$$t = t_0 + \epsilon, \quad \epsilon > 0. \quad (IV.3)$$

Similarly, it is now without ambiguity when we say that

the charge-current distribution is turned off at t_f . We shall also assume that ρ_{ex} and J_{ex} never become time-constant during any time interval of interest.

The complete α -Lorentz (EM) gauge is defined by

$$\nabla \cdot \mathbf{A}^{(\alpha L)} + \frac{1}{a^2 c^2} \frac{\partial}{\partial t} \phi^{(\alpha L)} = 0. \quad (IV.4)$$

(Here, the expression (EM) gauge is clear without further explanation.) The most general solutions to Eqs. (IV.1) and (IV.4) are

$$\phi^{(\alpha L)}(\mathbf{r}, t) = \phi_{ac}^{(\alpha L)}(\mathbf{r}, t) - \frac{1}{c} \frac{\partial}{\partial t} \chi^{(\alpha L)}(\mathbf{r}, t), \quad (IV.5)$$

$$\mathbf{A}^{(\alpha L)}(\mathbf{r}, t) = \mathbf{A}_c^{(L)}(\mathbf{r}, t) + c \int_{t_0}^t dt' [\nabla \phi_c^{(L)}(\mathbf{r}, t') - \nabla \phi_{ac}^{(\alpha L)}(\mathbf{r}, t')] + \nabla \chi^{(\alpha L)}(\mathbf{r}, t).$$

All the quantities in Eq. (IV.5) are defined below.

$$G(\mathbf{r}, t | \mathbf{b} | \mathbf{r}', t') = \frac{\delta(t' + \frac{|\mathbf{r} - \mathbf{r}'|}{b} - t)}{|\mathbf{r} - \mathbf{r}'|}, \quad b = c, ac \quad (IV.6)$$

$$\phi_{ac}^{(\alpha L)}(\mathbf{r}, t) = \int d\mathbf{r}' \int dt' G(\mathbf{r}, t | ac | \mathbf{r}', t') \rho_{ex}(\mathbf{r}', t'), \quad (IV.7)$$

$$\phi_c^{(L)}(\mathbf{r}, t) = \left[\phi_{ac}^{(\alpha L)}(\mathbf{r}, t) \right]_{\alpha=1}, \quad (IV.8)$$

$$\mathbf{A}_c^{(L)}(\mathbf{r}, t) = \frac{1}{c} \int d\mathbf{r}' \int dt' G(\mathbf{r}, t | c | \mathbf{r}', t') \mathbf{J}_{ex}(\mathbf{r}', t'), \quad (IV.9)$$

and $\chi^{(\alpha L)}(\mathbf{r}, t)$ is the solution of the differential equation

$$\left(\nabla^2 - \frac{1}{a^2 c^2} \frac{\partial^2}{\partial t^2} \right) \chi^{(\alpha L)}(\mathbf{r}, t) = 0. \quad (IV.10)$$

In order to discuss causality, we define a function $n(\mathbf{r}, t)$ by

$$n(\mathbf{r}, t) = 1 \quad (IV.11)$$

if not all of ρ_{ex} , $\mathbf{J}_{ex}$, $\nabla \rho_{ex}$, $\frac{\partial}{\partial t} \mathbf{J}_{ex}$, and $\nabla \times \mathbf{J}_{ex}$ vanish at $\mathbf{r}$ and t , and

$$n(\mathbf{r}, t) = 0 \quad (IV.12)$$

if otherwise. We then define a function $F_{ac}(\mathbf{r}, t)$ by

$$F_{ac}(\mathbf{r}, t) = \int d\mathbf{r}' \int dt' G(\mathbf{r}, t | ac | \mathbf{r}', t') n(\mathbf{r}', t'). \quad (IV.13)$$

From the definition of $F_{ac}(\mathbf{r}, t)$, we have,

$$F_c(\mathbf{r}, t) = 0 \Rightarrow E_c(\mathbf{r}, t) = B_c(\mathbf{r}, t) = 0. \quad (IV.14)$$

This function F_{ac} defines the so-called ac -front (in Fig. 1). An ac -front is defined to be the surface that separates the region where $F_{ac}(\mathbf{r}, t)$ does not vanish from the region where $F_{ac}(\mathbf{r}, t)$ does.

Now, we define the α -Lorentz gauge in quantum mechanics. We say that $A^{(\alpha L)}$, $\phi^{(\alpha L)}$ are potentials in the α -Lorentz gauge if

$$\begin{aligned} \phi^{(\alpha L)} &= \phi_{ac}^{(\alpha L)} - \frac{1}{c} \frac{\partial}{\partial t} f(t), \\ A^{(\alpha L)} &= A_c^{(L)} + c \int_{t_0}^t dt' [\nabla \phi_c^{(L)} - \nabla \phi_{ac}^{(\alpha L)}], \end{aligned} \quad (IV.15)$$

where $f(t)$ is any function satisfying the condition that its first order derivative exists and is continuous. Later in our discussions, we shall take $f(t) = 0$ since the exact form of this function does not affect probabilities (not probability amplitudes).

B. A Mathematical Result

Now, we must formulate a mathematical tool in order to connect our understanding of the propagation properties of the potentials in the α -Lorentz gauge with Lamb's criterion (i.e., to connect E.M. with Q.M.). Before we do

that, it is more convenient to recall some of the notations defined before and also to define some new quantities. First, we define

$$R(\Lambda) \equiv \exp\left[\frac{ie}{\hbar c} \Lambda(\mathbf{r}, t)\right], \quad (IV.16)$$

$$H_R^\Lambda(t) \equiv H(\Lambda = \nabla\Lambda, \phi = -\frac{1}{c} \frac{\partial\Lambda}{\partial t}) - H_B^0.$$

As clearly indicated in Eq. (IV.16), H_R^Λ is the "interaction" Hamiltonian when the vector potential is $\nabla\Lambda$ and the scalar potential is $-\frac{1}{c} \frac{\partial\Lambda}{\partial t}$. We then recall that $\{\epsilon_1\}$ are the eigenvalues of the "unperturbed" Hamiltonian H_B^0 and $\{\phi_1(\mathbf{r})\}$ are the orthonormal and complete (assumed) eigenfunctions of H_B^0 . Now we formulate our result below.

Theorem 1: If $\Lambda(\mathbf{r}, t)$ is a gauge function (see Sect. II.B) and if $\nabla\Lambda(\mathbf{r}, t)$ does not vanish identically in the region and during the time interval of concern, then, for every given i , there exists at least a $j \neq i$ with $\epsilon_j \neq \epsilon_i$ such that

$$\langle \phi_j | R(\Lambda) | \phi_i \rangle \neq 0,$$

and

$$\langle \phi_j | H_R^\Lambda(t) | \phi_i \rangle \neq 0.$$

(Note, here by $\neq 0$ we do not mean that the quantity of interest must be big. For example, if the quantity has absolute value of the order, say, 10^{-100} , it is still considered as $\neq 0$ here.)

Proof: First, we note that $\nabla\Lambda \neq 0$ guarantees that

$$[H_B^0, R(\Lambda)] \neq 0 \quad (IV.17)$$

and

$$[H_B^0, H_R^\Lambda] \neq 0.$$

Because of the above commutator relations, the eigenfunctions of H_B^0 can not be simultaneously the eigen-

functions of both $R(\Lambda)$ and H_R^Λ . Therefore, our theorem is obviously true for non-degenerate situations. A slightly more careful consideration also tells us that our argument is also true for degenerate cases. This completes our proof.

C. "Wrong" Gauges

Now, we have prepared ourselves well enough for the main theme of this paper: to locate those gauges in which the theoretical probabilities constructed according to Eq. (II.8) can not be the physically true probabilities. As has been repeatedly emphasized before, this objective will be accomplished only through Lamb's criterion -- with our substitute for the physically true probabilities by the requirements from relativity and the conservation laws. To be more specific, we define that a "change of the state"¹⁴ violates causality if this change takes place in a region where $F_c(\mathbf{r}, t)$ vanishes identically at the time instant of concern. Such a change is said to violate the conservation laws if it takes place in a region where $F_c(\mathbf{r}, t)$ does not vanish but $E_c(\mathbf{r}, t)$ and $B_c(\mathbf{r}, t)$ do vanish. We also mention that only the complete α -Lorentz gauge will be investigated here.

For the sake of simplicity and preciseness without loss of generality, we assume that there exists a finite (and generally small) region denoted by S_Q such that

$$P(\mathbf{r}, t) = 0 \quad \text{for all } \mathbf{r} \notin S_Q \text{ and for all } t, \quad (IV.18)$$

where $P(\mathbf{r}, t)$ is the probability of finding the particle at position $\mathbf{r}$ and at time t . The location of S_Q will be varied to suit our purpose. Lastly, we shall also assume that our quantum particle is in the ground state.

Now, we investigate the α -Lorentz gauge with $\alpha > 1$. We first notice that for every $t > t_0$ (for the meaning of t_0 see Sect. IV.A.) there exists a non-empty region $S_{c,\alpha c}(t)$ such that inside $S_{c,\alpha c}(t)$, $\phi_{ac}^{(\alpha L)}$ does not vanish, however small it may be, if the external charge density involved is arbitrary. How big this region $S_{c,\alpha c}(t)$ is depends upon the value of t for each chosen α (See Fig. 1).

In order to apply Theorem I, we first note that $A_{ac}^{(\alpha L)}, \phi_{ac}^{(\alpha L)}$ can be decomposed into two distinctive components, for $\alpha \neq 1$,

$$(A_{ac}^{(\alpha L)}, \phi_{ac}^{(\alpha L)}) = (A_{ac}^{(\alpha L)}, 0) + (A_{ac}^{(\alpha L)}, \phi_{ac}^{(\alpha L)}) \quad (IV.19)$$

where $A_{ac}^{(\alpha L)}$ and $A_{ac}^{(\alpha L)}$ can be directly identified from Eq. (IV.15) by propagation indices. Furthermore, we notice that

$$A_{ac}^{(\alpha L)} = \nabla A_{ac}^{(\alpha L)}, \quad \phi_{ac}^{(\alpha L)} = -\frac{1}{c} \frac{\partial}{\partial t} A_{ac}^{(\alpha L)} \quad (IV.20)$$

where

$$A_{ac}^{(\alpha L)}(r, t) = -c \int_{t_0}^t dt' \phi_{ac}^{(\alpha L)}(r, t') \quad (IV.21)$$

By assumption, $A_{ac}^{(\alpha L)}(r, t)$ is a gauge function.

Now, we wait until the "thickness" of the region $S_{c,\alpha c}(t)$ is large enough so that the region S_Q can be completely contained inside $S_{c,\alpha c}(t)$. If we locate S_Q inside $S_{c,\alpha c}(t)$ at time t , then the wave function at time t is

$$\psi^{(\alpha L)}(r, t) = b R(A_{ac}^{(\alpha L)}) e^{-i E_{gs} t / \hbar} \phi_{gs}(r), \quad (IV.22)$$

where b is a constant of modulus one and

$$R(A_{ac}^{(\alpha L)}) = \exp\left[\frac{ie}{\hbar c} A_{ac}^{(\alpha L)}(r, t)\right].$$

Now, "the usual expansion coefficients" $\{C_1^{(\alpha L)}(t)\}$ are, according to Eq. (II.5),

$$C_1^{(\alpha L)}(t) = \langle \phi_1 | R(A_{ac}^{(\alpha L)}) | \phi_{gs} \rangle b e^{-i E_{gs} t / \hbar} \quad (IV.23)$$

By Theorem I, we know that the interpretation of $|C_1^{(\alpha L)}(t)|^2$ as the probability of finding the particle in the state with energy ϵ_1 violates causality. This means that "the usual interpretation" in the α -Lorentz gauge with $\alpha > 1$ violates causality. Hence, in our language, the complete α -Lorentz gauge with $\alpha > 1$ is a "wrong" gauge.

Similarly, one can show that the α -Lorentz gauge with $0 < \alpha < 1$ is also a "wrong" gauge. (This, however, requires that the sources ρ_{ex} and J_{ex} be turned off at some finite time.)

From these arguments, we are led to considering only those gauges that have potentials propagating at speed c only. An obvious choice of our next investigation is the Lorentz gauge. This is primarily because the potentials in the Lorentz gauge satisfy all the requirements of (special) relativity. This, however, does not mean that the Lorentz gauge entails no difficulties at all. The most well-known difficulty associated with this gauge is the longitudinal component. In accordance to our spirit, we also raise a special example to demonstrate how this longitudinal component can cause undesirable consequences in our conventional interpretation (of quantum mechanics).

We now turn our attention to Fig. 2. In Fig. 2, we have a line source (ρ_{ex} and $\hat{z} J_{ex}$) with center at the origin of the $x'y'z'$ -axes and the line is on the z' -axis. We also assume that the length of this source is very small and that, for the simplicity of arguments, the source has a definite frequency ω . Under this situation, the potentials in the Lorentz gauge are

$$\begin{aligned} A_c^{(L)}(r', t') &= \frac{1}{c} \hat{z} \int d\tau'' \int dt'' G(r', t' | c | r'', t'') J_{ex}(r'', t'') \\ \phi_c^{(L)}(r', t') &= \int d\tau'' \int dt'' G(r', t' | c | r'', t'') \rho_{ex}(r'', t'') \end{aligned} \quad (IV.24)$$

Now we place our quantum region S_Q on the z' -axis "very far" away from the origin. If we assume that S_Q is "very small", then to the first order approximation (dipole approximation), all electromagnetic fields vanish identically inside S_Q . This means that $A_c^{(L)}$ and $\phi_c^{(L)}$ in Eq. (IV.24) can be approximated by

$$\begin{aligned} A_c^{(L)}(\mathbf{r}', t') &= \hat{z} [a_3 e^{ikz' - i\omega t'} + a_3^* e^{-ikz' + i\omega t'}] \\ \phi_c^{(L)}(\mathbf{r}', t') &= [a_0 e^{ikz' - i\omega t'} + a_0^* e^{-ikz' + i\omega t'}] \end{aligned} \quad (IV.25)$$

within our quantum system S_Q , where $k = \omega/c$. Here, it should be noted that although the local fields inside S_Q can not determine uniquely a_3 and a_0 , the values of a_3 and a_0 are completely and uniquely determined by ρ_{ex} and J_{ex} and the distance between the origin and the quantum system.¹⁵ The values of a_0 and a_3 satisfy the Lorentz condition

$$\nabla' \cdot A_c^{(L)} + \frac{1}{c} \frac{\partial}{\partial t'} \phi_c^{(L)} = 0$$

It is also understood that Eq. (IV.25) is valid only when F_c does not vanish inside S_Q .

Now, we re-write Eq. (IV.25) in the variables $xyzt$ of our quantum system. To make life simple, we assume that xyz has the same orientation as $x'y'z'$ and that xyz is at rest with respect to $x'y'z'$. Thus, we have

$$\begin{aligned} A_c^{(L)}(\mathbf{r}, t) &= \hat{z} [A_3 e^{ikz - i\omega t} + A_3^* e^{-ikz + i\omega t}] , \\ \phi_c^{(L)}(\mathbf{r}, t) &= [A_0 e^{ikz - i\omega t} + A_0^* e^{-ikz + i\omega t}] , \end{aligned} \quad (IV.26)$$

where A_3 and A_0 have absorbed all the necessary phase factors due to the change of variables. In order to apply Theorem 1, we notice that

$$A_c^{(L)} = \nabla A_c^{(L)}, \quad \phi_c^{(L)} = -\frac{1}{c} \frac{\partial}{\partial t} A_c^{(L)}. \quad (IV.27)$$

with

$$A_c^{(L)} = -c \int_{t_0}^t dt' \phi_c^{(L)}(\mathbf{r}, t'). \quad (IV.28)$$

Then by the same arguments used to exclude the α -Lorentz gauge, we can conclude that under our special situation the Lorentz gauge is also a "wrong" gauge because the conservation laws are violated.

At this moment, one is tempted to try to eliminate this unpleasant longitudinal component. (One way to do this has been described in Ref. [2].)

But before embarking on such an adventure, we must point out that as long as we interpret $\frac{1}{8\pi}(\mathbf{E}^2 + \mathbf{B}^2)$ as the local field energy density, the only way to avoid committing one mistake after another is to investigate the possibility of the existence of a gauge with potentials A', ϕ' such that A' and ϕ' vanish wherever $\mathbf{E}$ and $\mathbf{B}$ vanish. Furthermore, A' and ϕ' must be able to generate the total fields $\mathbf{E}$ and $\mathbf{B}$ everywhere and must also not contain terms that violate the requirements from relativity. One of these conditions is that A' and ϕ' must both propagate at speed c . Whether it is possible to construct such a gauge will not be resolved here. Although it is clear that such a set of potentials exists in the radiation zone [2], it is not at all clear whether this gauge exists in other regions. This will be further discussed in Section V.B.

V. SUMMARY AND DISCUSSIONS

A. Summary

This paper deals with the question whether the conventional interpretation of probability amplitudes is gauge invariant. We only work within the framework of treating electromagnetic fields and potentials as classical and external quantities. In order to understand the reasons why "the usual interpretation" is, or is not, gauge invariant, the first step is to go back to re-examine very carefully the whole procedure of constructing the so-called "usual" probability amplitudes in two arbitrary, different gauges. (We always stress that the usage of the term gauge in this paper is not the same as that in classical electromagnetic theory.) Here, we must also point out that we are not dealing with perturbation theory. Every result in this paper is exact as far as the interpretation of probability amplitudes is concerned.

Upon examination of the procedures of constructing the probability amplitudes in two different gauges, we have found that identical procedures must not be used in any two different gauges if consistent interpretations in these two different gauges are desired. However, if consistent interpretations are demanded, different procedures for constructing probability amplitudes must be used for different gauges.

Here, we must clarify several important points concerning the above-mentioned consistent interpretations and different procedures for different gauges. Here, consistent interpretations only mean that the (theoretical) probabilities constructed according to the designated procedures are mathematically consistent with one another. At no time does this imply that the theoretical probabilities conform to physical reality, nor does it mean that

these probabilities are gauge invariant. This is because when different procedures must be used for different gauges, the gauge in which the probability amplitudes are to be defined "in the usual way" can not be arbitrary. It therefore must be a "special" gauge.

How to choose (or more properly, find) this "special" gauge becomes the most pertinent issue now. It is because if a "wrong" gauge is chosen to start with, all the subsequent interpretations (in different gauges) that are consistent with the interpretation in this "wrong" gauge will be consistently wrong. This explains why care must be exercised and the underlying physics must be as convincing as possible in selecting this "special" (starting) gauge.

So, it becomes absolutely necessary to examine the underlying physics in choosing this "special" starting gauge. Here, we find that the emphasis of the "traditional" reasoning is on the physical correctness of the "interaction" Hamiltonians. We must be aware that there is nothing wrong with this logic if there is some method devised to test, in a mathematically rigorous way, the physical correctness of any "interaction" Hamiltonian. This is why we believe that the procedure used by Lamb in arriving at his conclusion concerning $A \cdot p$ vs. $E \cdot r$ is much, much more superior to any of the "traditional" arguments. This procedure is later formalized into Lamb's criterion which becomes the most important conceptual guidance of this paper.

How Lamb's criterion is applied characterizes the differences between Lamb's and our approaches to the gauge problem. Armed with the observed probabilities from experiments, Lamb was able to establish that $E \cdot r$ is the only correct "interaction" Hamiltonian. This, however, we can not do since we do not have the knowledge of the physically true probabilities in general. Instead, our approach consists of demanding that the conservation laws and

the restrictions from relativity on the propagation of electromagnetic physical information not be violated. Note, our statements concerning the conservation laws are correct only if we accept the physical interpretation of the Poynting Theorem in classical electrodynamics [8]. That is, our approach only allows us to determine "wrong" gauges. With the help from our investigation of the propagation characteristics of the complete α -Lorentz gauge, we conclude that "the usual interpretation" in the α -Lorentz gauge violates causality when $\alpha \neq 1$. Furthermore, we also give an example to show that "the usual interpretation" in the Lorentz gauge violates the conservation laws.

B. Discussions

The most important concept in this paper is Lamb's criterion. It allows that the correctness of "the usual matrix elements" of the "interaction" Hamiltonian in an arbitrary gauge be determined by the correctness of the "usual" probabilities. From our investigation, we have learned that the only gauge that need be seriously considered is the gauge in which the potentials A and ϕ satisfy the following electromagnetic gauge condition:

$$\nabla \cdot \underline{A} + \frac{1}{c} \frac{\partial \phi}{\partial t} = - \frac{4\pi}{c} f(\underline{r}, t). \quad (V.1)$$

Here the function $f(\underline{r}, t)$ must satisfy all the conditions that are to be derived (not assumed) from all the reasonable restrictions that the physically true probabilities are required to satisfy.

Our previous experiences (Section IV.C) tell us that consideration of causality demands that $f(\underline{r}, t)$ vanish at those $\underline{r}$ and t where both $\rho_{\text{ex}}(\underline{r}, t)$ and $\underline{J}_{\text{ex}}(\underline{r}, t)$ vanish so that the resulting potentials always propagate at speed c from the source region. Consideration of the conservation laws leads to the result that $f(\underline{r}, t)$ must not vanish identically since this would lead to the

Lorentz gauge. Furthermore, it also requires that $f(\underline{r}, t)$ be such that the resulting potentials vanish at any $\underline{r}$ and t where both $\underline{E}_c(\underline{r}, t)$ and $\underline{B}_c(\underline{r}, t)$ vanish. One consequence of this is that A and ϕ must reduce in the radiation zone to the potentials in the radiation gauge $A^{(R)}$ and $\phi^{(R)}$ defined in Section IV of Ref. [2]. Another further restriction on $f(\underline{r}, t)$ is that it must not have dependence on $\rho_{\text{ex}}(\underline{r}, t')$ and $\underline{J}_{\text{ex}}(\underline{r}, t')$ at $t' \neq t$. The reason is that such a dependence will be passed on to A and ϕ , then to "the usual matrix elements" of the "interaction" Hamiltonian in this gauge, and finally to the "usual" probabilities in this gauge. Hereafter, this gauge will be referred to as the "good" gauge. Note, the Lorentz invariance of $f(\underline{r}, t)$ is not required here.

Whether or not the "good" gauge exists is of vital importance to the validity and also the convincing ability of "the usual interpretation of probability amplitudes". There is a priori no reason to presume that such a gauge exists just because the Schrödinger equation is valid in any arbitrary gauge. This is because the Schrödinger equation (in any gauge) governs only the time-evolution of the wave function (in the same gauge) whereas "the usual probabilities" (in this gauge) are defined by the wave function and the "usual" basis functions. These "usual" basis functions are not within the control of the Schrödinger equation since they are obtained without any direct and unique reference to the Hamiltonian or the Schrödinger equation in any arbitrary gauge.¹⁶ This explains why the formal gauge invariance of the Schrödinger equation can not and does not guarantee that the "usual" probabilities (in a specific gauge) are gauge invariant and hence that these "usual" probabilities, which are observables by assumption within this "usual" interpretive scheme, will automatically satisfy all necessary conditions of an observable. From this understanding, it is then clear that if our

"good" gauge exists, it must be shown explicitly. This also explains why the procedure of first choosing a gauge and then constructing the basis functions according to the equations of motion of operators in the same gauge [1] is in general incompatible with the procedure of first choosing the "usual" basis functions as our basis functions and then trying to find a gauge for the Schrödinger equation.¹⁷

Now, we must point out that here we are dealing with only one of the many technical difficulties of "the usual interpretation" of quantum mechanics. This can be easily understood because we only investigate the consequences of constructing the probability amplitudes "in the usual way" (i.e., by Eq.(II.5)) and of physically interpreting these probability amplitudes by Eq. (II.8). The fundamental concept that leads to this "usual" interpretation is the decomposition of a Hamiltonian into the "unperturbed" Hamiltonian and an "interaction" Hamiltonian and the physical interpretation of the "unperturbed" Hamiltonian as the particle's energy operator. Our investigation of "the usual interpretation" strongly implies that it is generally wrong, unless on some very restrictive occasions, to interpret the "unperturbed" Hamiltonian as the particle's energy operator. This result can be obtained by studying the Hamiltonian formalism of classical mechanics where the observability of all classical "unperturbed" Hamiltonians¹⁸ can be unambiguously determined in all situations. This explains why we do not try to resolve the question of existence of our "good" gauge in Eq. (V.1).

ACKNOWLEDGEMENTS

I would like to thank Professor Fried ([5]) for sending me the reprint of his paper and Professor Swift ([6]) for explaining to me the details of how he could contradict Lamb's claim on A. p.

FOOTNOTES

¹ The expression "the usual interpretation of probability amplitudes", which often will be shortened to "the usual interpretation", is a quotation from Ref. [3]. We shall always retain the quotation marks to emphasize that this expression means exactly what is meant by Lamb, which is to define probability amplitudes by the eigenfunctions of the "unperturbed" Hamiltonian.

² It should be noted that Refs. [5] and [6] dispute only Lamb's claim on $\mathbf{A} \cdot \mathbf{p}$. Lamb's claim on $\mathbf{E} \cdot \mathbf{r}$ is neither explained nor challenged by these two authors.

³ I am grateful to Prof. W. R. Bennett, Jr., for a short discussion on this point.

⁴ For both relativistic and nonrelativistic quantum mechanics, our definition for $\langle \phi_i | A | \phi_j \rangle$ is

$$\langle \phi_i | A | \phi_j \rangle \equiv \int d\mathbf{r} \phi_i^* A \phi_j.$$

⁵ Whenever the word "usual" appears, it always means that the basis functions involved are the eigenfunctions of the "unperturbed" Hamiltonian. These eigenfunctions are also termed the "usual" basis functions.

⁶ The definition in Eq. (II.14) is not necessary for our gauge invariant interpretation in Ref. [1].

⁷ One example of a gauge condition, which by our usage is used exclusively in classical electrodynamics, is the Lorentz (gauge) condition

$$\nabla \cdot \mathbf{A} + \frac{1}{c} \frac{\partial \phi}{\partial t} = 0.$$

Note the difference between this definition and Eq. (II.14).

⁸ Notice that the interpretation in Eq. (II.21) is not "the usual interpretation" because the basis functions are not the "usual" basis functions. Compare this with Eqs. (II.5) and (II.19), and also observe that there are two superscripts in Eq. (II.21) instead of only one in Eqs. (II.5) and (II.19).

9. Any argument not based on Lamb's criterion to be formulated in Sect. III.B will be classified as a "traditional" argument. We also mention that no reference to any "traditional" argument will be quoted here.
10. One should be aware that the word transverse is subject to two non-identical interpretations. The first one is divergenceless as in the "transverse" gauge ($\nabla \cdot \mathbf{A} = 0$). The second interpretation is perpendicular to the direction of propagation from the source region as in the expression the "acceleration" fields of a point charged particle are transverse [7].
11. Before one uses Lamb's criterion to judge the correctness of the "usual" probabilities (in a gauge) by the correctness of the "interaction" Hamiltonian (in the same gauge), one must first prove that this "interaction" Hamiltonian is observable since it is meaningless to say that a non-observable is physically correct or wrong or is physically preferred or not.
12. The gauge invariant interpretation in Ref. [1] offers an example of this viewpoint.
13. See Eq. (4.33) of Ref. [1].
14. We offer no formal definition of the expression "change of the state." What we mean by this expression can be understood from our examples to be discussed in this section.
15. Although from fields' point of view, transforming away the longitudinal component of the vector potential in the Lorentz gauge, which can be done only in the radiation zone according to our understanding [2], changes nothing, this procedure reduces the original charge and current densities to effective (transverse) current density from potentials' (in the Lorentz gauge only) point of view. See Sect. IV of Ref. [2].

16. The interpretation in Eq. (II.21) does not refer uniquely to gauge g' . This is because χ refers to the difference of two gauges.
17. This has been very carefully discussed in Ref. [1], especially Sect. III.
18. In classical mechanics and in Heisenberg's representation, the number of "unperturbed" Hamiltonians is (uncountably) infinite. This is because the number of gauges is also (uncountably) infinite and because a canonical momentum can be uniquely defined only after a gauge has been specified.

REFERENCES

- [1] K.-H. Yang, Univ. of Md. Tech. Rpt. #76-058, to be published in Annals of Physics (N.Y.).
- [2] K.-H. Yang, Univ. of Md. Tech. Rpt. #76-079, to be published in Annals of Physics (N.Y.).
- [3] W. E. Lamb, Jr., Phys. Rev. 85, 259 (1952). Also see W. E. Lamb, Jr., and R. C. Retherford, Phys. Rev. 79, 549 (1950); 81, 222 (1951).
- [4] E. A. Power and S. Zienan, Philos. Trans. R. Soc. (London) 251, 427 (1959).
- [5] Z. Fried, Phys. Rev. A8, 2835 (1973). (Also see Bull. Am. Phys. Soc. Ser. II, 18, 122 (1973).)
- [6] A. R. Swift (private communication, 1971-1972, unpublished).
- [7] See, for example, J. D. Jackson, "Classical Electrodynamics," John Wiley, New York, 1962, Sects. 14.1-14.4.
- [8] See, for example, Sects. 6.8 and 6.9 of Ref. [7].

FIGURE CAPTIONS

Fig. 1: The set-up of a hypothetical experiment designed to exclude the use of the complete α -Lorentz gauge with $\alpha > 1$ in the Schrödinger equation when probability amplitudes are defined "in the usual way."
(See Sects. IV.A and IV.C.)

Fig. 2: Same as in Fig. 1 except for the Lorentz gauge.

Fig. 1

Fig. 2