ORO 5126-23

GAUGE INVARIANCE OF THE RAYLEIGH-SCHRÖDINGER
TIME-INDEPENDENT PERTURBATION THEORY

Kuo-Ho Yang

Department of Physics and Astronomy
University of Maryland, College Park, Maryland 20742

August 1977

U. of Md. TR #78-012
U. of Md. PP #78-036

MASTER

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

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

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.

ORO 5126-23
U. of Md. TR #78-012
U. of Md. PP #78-036

GAUGE INVARIANCE OF THE RAYLEIGH-SCHRÖDINGER
TIME-INDEPENDENT PERTURBATION THEORY

Kuo-Ho Yang[†]

Department of Physics and Astronomy
University of Maryland, College Park, Maryland 20742

August 1977

ABSTRACT

We show that the Rayleigh-Schrödinger time-independent perturbation theory is gauge invariant when the operator concerned is the particle's instantaneous energy operator

$$H_B = \frac{1}{2m} \left(\vec{p} - \frac{e}{c} \vec{A} \right)^2 + eV_0.$$

More explicitly, we show that the energy perturbation corrections of each individual order of every state is gauge invariant. When the vector potential is curless, the energy corrections of all orders are shown to vanish identically regardless of the explicit form of the vector potential. The relationship between causality and gauge invariance is investigated. We show that gauge invariance guarantees conformity with causality and violation of gauge invariance implies violation of causality.

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, 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

Gauge invariance requires that whatever set of potentials describing a given electromagnetic field is used, identical physical results be obtained through identical procedures of computations. Implicit in the requirement is that any set of potentials is no more special than any other. That is, whatever gauge is used is legitimate. Note, throughout this paper, two gauges are defined to be different if the vector potentials are different.¹

It is this fact that any gauge is allowed that has been widely exploited. In literature, one commonly used maneuver consists of starting from a certain gauge often purported to be arbitrary, then transforming to another gauge by a gauge transformation, and finally formulating the problem in this newly acquired gauge.² Often, although not always, some explanations are given to justify the necessity of working on the new gauge. They are nothing more than that the gauge transformation involved has some special physical properties and that the gauge after the transformation possesses desirable physical and mathematical properties that are absent in the gauge before the transformation.²⁻⁶

Whether or not this procedure of going from an arbitrary gauge to a special and desirable gauge actually produces a gauge invariant formulation of quantum mechanics has long been a controversy.³⁻²² This we have investigated carefully in previous papers²³ and therefore will not be discussed here. In this paper, we shall be concerned only with time-independent perturbation theory²⁴⁻²⁸ (TIPT). Here, by TIPT we mean either time is not involved or if it is it appears as a parameter rather than as a variable.

The immediate reason for our interest in TIPT stems from the fact that the gauge invariant²⁹ instantaneous energy operator introduced in Ref. 19, also known as the basis-defining Hamiltonian H_B defined by

$$H_B = \frac{1}{2m} \left(p - \frac{e}{c} A \right)^2 + eV_0 \quad (1.1)$$

for a spinless, nonrelativistic charged particle, shows explicit dependence on the vector potential while possesses a gauge invariant energy spectrum. Since in most practical situations, approximation methods are the only means available, the most crucial question is whether a perturbation method will give gauge invariant results despite the explicit gauge dependence of the energy operator. This is because if all orders of the energy perturbation corrections are to be physically meaningful, each individual term must be gauge invariant.

The purpose of this paper is to show analytically that the classic Rayleigh-Schrödinger time-independent perturbation theory³⁰ (RSTIPT) is gauge invariant when the operators considered are the basis-defining Hamiltonians in Eq. (1.1). More explicitly, we are going to prove that the eigenvalue perturbation corrections for two basis-defining Hamiltonians related to each other by a gauge transformation are identical--order by order. Our insistence of gauge invariance of all orders of energy corrections originates from our conviction that only gauge invariant quantities can have physical meanings. (This, however, does not imply that all gauge invariant quantities have physical meanings.)

We organize this paper in the following manner. We briefly review in Sect. II the RSTIPT for non-degenerate cases. After the review, it will be shown that the orthogonality constraint usually imposed upon the eigenfunction corrections is really not necessary. (This will be related eventually to gauge invariance.) In Sect. III, we show that when the vector potential in Eq. (1.1) is curless, the zeroth order eigenvalues will be the exact ones while all orders of corrections will vanish identically. This result is shown to be in conformity with causality. In Sect. IV, we show that even

when the vector potential represents a non-vanishing magnetic field, the energy corrections will still be gauge invariant. Again, the relationship between gauge invariance and causality is investigated; we find that gauge invariance guarantees that causality is not violated. Finally, conclusions will be presented in Sect. V.

II. THE RAYLEIGH-SCHRÖDINGER TIME-INDEPENDENT PERTURBATION THEORY

A. A Brief Review

The Rayleigh-Schrödinger time-independent perturbation theory for non-degenerate cases will be briefly reviewed here. We shall be as succinct as possible since the basic ideas and techniques are available in every textbook on elementary quantum mechanics.³⁰ However, since what will interest us requires a form slightly different from that in textbooks, we will have to devote some space for the basic formalism.

Assume we have a Hermitian operator³¹ $H(\lambda)$, where λ is a parameter. The objective is to find the eigenvalues and the eigenfunctions of this operator. Let us also assume that the eigenfunctions $\{\psi_i^{(0)}\}$ of the operator H_0 are known and form a complete set, where H_0 is defined by

$$H_0 \equiv H(\lambda) \Big|_{\lambda=0} \quad (2.1)$$

so that we can expand the eigenfunctions of $H(\lambda)$, for all λ , in the Hilbert space spanned by $\{\psi_i^{(0)}\}$. Let us now denote the eigenvalues of H_0 by $\{E_i^{(0)}\}$; then

$$H_0 \psi_i^{(0)} = E_i^{(0)} \psi_i^{(0)}. \quad (2.2)$$

For convenience, $\{\psi_i^{(0)}\}$ are ortho-normalized.

$$\langle \psi_i^{(0)} | \psi_j^{(0)} \rangle = \delta_{ij} \quad (2.3)$$

We must call to the attention that throughout this paper, only vectors in Hilbert space are concerned. Those that are not squarely normalizable are not within our consideration.

Let us use $\{E_i(\lambda)\}$ and $\{\psi_i(\lambda)\}$ for the eigenvalues and the eigenfunctions of $H(\lambda)$, i.e.,

$$H(\lambda)\psi_i(\lambda) = E_i(\lambda)\psi_i(\lambda). \quad (2.4)$$

Following traditional procedure, we assume that $H(\lambda)$, $E_i(\lambda)$, and $\psi_i(\lambda)$ can be expanded in Taylor's series.

$$H(\lambda) = \sum_{n=0} \lambda^n H^{(n)}, \quad (2.5)$$

$$E_i(\lambda) = \sum_{n=0} \lambda^n E_i^{(n)}, \quad (2.6)$$

and

$$\psi_i(\lambda) = \sum_{n=0} \lambda^n \psi_i^{(n)}. \quad (2.7)$$

For our specific purpose, the series in Eq. (2.5) is to terminate after the third term. That is

$$H(\lambda) = H_0 + \lambda V_1 + \lambda^2 V_2 \quad (2.8)$$

where we have used H_0 , V_1 , and V_2 for $H^{(0)}$, $H^{(1)}$, and $H^{(2)}$, respectively.

Substituting Eqs. (2.6) to (2.8) into Eq. (2.4), and then equating the coefficients on both sides of the same power of λ , we get

$$H_0 \psi_i^{(0)} = E_i^{(0)} \psi_i^{(0)} \quad (2.9.0)$$

$$H_0 \psi_i^{(1)} + V_1 \psi_i^{(0)} = E_i^{(0)} \psi_i^{(1)} + E_i^{(1)} \psi_i^{(0)}, \quad (2.9.1)$$

$$H_0 \psi_i^{(2)} + V_1 \psi_i^{(1)} + V_2 \psi_i^{(0)} = E_i^{(0)} \psi_i^{(2)} + E_i^{(1)} \psi_i^{(1)} + E_i^{(2)} \psi_i^{(0)} \quad (2.9.2)$$

and

$$H_0 \psi_i^{(n)} + V_1 \psi_i^{(n-1)} + V_2 \psi_i^{(n-2)} = \sum_{j=0}^n E_i^{(j)} \psi_i^{(n-j)}. \quad (2.9.n)$$

It is customary to impose the orthogonality constraint³⁰

$$\langle \psi_i^{(0)} | \psi_i^{(n)} \rangle = 0, \quad n \geq 1. \quad (2.10)$$

Using Eqs. (2.9) and (2.10), we have the expressions for the energy corrections:

$$E_i^{(1)} = \langle \psi_i^{(0)} | V_1 | \psi_i^{(0)} \rangle \quad (2.11.1)$$

$$E_i^{(n)} = \langle \psi_i^{(0)} | V_1 | \psi_i^{(n-1)} \rangle + \langle \psi_i^{(0)} | V_2 | \psi_i^{(n-2)} \rangle - \sum_{j=1}^{n-1} E_i^{(j)} \langle \psi_i^{(0)} | \psi_i^{(n-j)} \rangle. \quad (2.11.n)$$

Note, in Eq. (2.11.n), the last term vanishes identically because of Eq. (2.10).

The reason that this term is kept here is because, as we shall show later, the orthogonality constraint Eq. (2.10) is really not necessary. When Eq. (2.10) is not imposed, this summation becomes necessary.

The standard procedure in this perturbation is to first solve for $E_i^{(1)}$, then for $\psi_i^{(1)}$, then $E_i^{(2)}$, $\psi_i^{(2)}$, etc. Although such a procedure is tedious, in theory at least, it is possible to compute the spectrum by obtaining successively the various orders of corrections for the energies.

Now we must stress that our goal is not to examine the convergence properties of this perturbative scheme. The one thing we want to investigate is whether the energy corrections $\{E_i^{(n)}\}$ given by Eq. (2.11) can have any physical significance when they do converge. In order that these corrections have any physical meanings, two necessary conditions, among many others, must be satisfied. They are: These corrections be independent of the orthogonality constraint Eq. (2.10) and they be gauge invariant. (We shall show in Sect. IV that these two conditions are actually one for operators of the type of Eq. (1.1).) First, we prove that Eq. (2.10) is not a necessity but only a convenience.

B. Independence of the Orthogonality Constraint

In order to show that Eq. (2.10) is really not necessary, we first assume that the quantities $\{u_i^{(n)}\}$ and $\{\epsilon_i^{(n)}\}$ satisfy the following equations.

$$H_0 u_i^{(0)} = \epsilon_i^{(0)} u_i^{(0)} \quad (2.12.0)$$

$$H_0 u_i^{(1)} + V_1 u_i^{(0)} = \epsilon_i^{(0)} u_i^{(1)} + \epsilon_i^{(1)} u_i^{(0)} \quad (2.12.1)$$

$$H_0 u_i^{(n)} + V_1 u_i^{(n-1)} + V_2 u_i^{(n-2)} = \sum_{j=0}^n \epsilon_i^{(j)} u_i^{(n-j)} \quad (2.12.n)$$

and

$$\epsilon_i^{(1)} = \langle u_i^{(0)} | V_1 | u_i^{(0)} \rangle \quad (2.13.1)$$

$$\epsilon_i^{(n)} = \langle u_i^{(0)} | V_1 | u_i^{(n-1)} \rangle + \langle u_i^{(0)} | V_2 | u_i^{(n-2)} \rangle - \sum_{j=1}^{n-1} \epsilon_i^{(j)} \langle u_i^{(0)} | u_i^{(n-j)} \rangle. \quad (2.13.n)$$

Since we do not wish to impose the orthogonality constraint on $u_i^{(n)}$ for $n \geq 1$, we define the quantities $\{\delta_i^{(n)}\}$ by

$$\delta_i^{(n)} \equiv \langle u_i^{(0)} | u_i^{(n)} \rangle. \quad (2.14)$$

The value each $\delta_i^{(n)}$ can assume can be arbitrary for n greater than 0. When n is zero, $\delta_i^{(0)}$ must be one.

Now we will show that $\{\epsilon_i^{(n)}\}$ given by Eq. (2.13) are identical to $\{E_i^{(n)}\}$ by Eq. (2.11).

First, we observe that $\{\epsilon_i^{(0)}\}$ and $\{E_i^{(0)}\}$ are identical since both are the spectrum of H_0 . So we have

$$\epsilon_i^{(0)} = E_i^{(0)}, \quad (2.15.0)$$

$$u_i^{(0)} = \psi_i^{(0)}. \quad (2.16.0)$$

Equation (2.16.0) immediately leads to

$$\epsilon_i^{(1)} = E_i^{(1)} \quad (2.15.1)$$

because of Eqs. (2.11.1) and (2.13.1). Utilizing all of the above results into Eq. (2.12.1) we get

$$(H_0 - E_i^{(0)}) u_i^{(1)} = (E_i^{(1)} - V_1) \psi_i^{(0)}.$$

If Eq. (2.9.1) is subtracted from the above equation, we have

$$(H_0 - E_i^{(0)}) (u_i^{(1)} - \psi_i^{(1)}) = 0$$

which immediately gives

$$u_i^{(1)} = \psi_i^{(1)} + \delta_i^{(1)} \psi_i^{(0)}. \quad (2.16.1)$$

Using all the results of Eqs. (2.15.0), (2.15.1), (2.16.0), and (2.16.1) into the expression for $\epsilon_i^{(2)}$, we can show the equality of $\epsilon_i^{(2)}$ and $E_i^{(2)}$ as follows.

$$\begin{aligned} \epsilon_i^{(2)} &= \langle u_i^{(0)} | V_1 - \epsilon_i^{(1)} | u_i^{(1)} \rangle + \langle u_i^{(0)} | V_2 | u_i^{(0)} \rangle \\ &= \langle \psi_i^{(0)} | V_1 | \psi_i^{(1)} \rangle + \langle \psi_i^{(0)} | V_2 | \psi_i^{(0)} \rangle + \delta_i^{(1)} \langle \psi_i^{(0)} | V_1 - E_i^{(1)} | \psi_i^{(0)} \rangle \\ &= E_i^{(2)} \end{aligned} \quad (2.15.2)$$

since the last term vanishes due to Eq. (2.11.1).

Now we want to find the relation between $u_i^{(2)}$ and $\psi_i^{(2)}$. The way to proceed is as follows. We use all previous results in the equation for $u_i^{(2)}$ to get

$$(H_0 - E_i^{(0)}) u_i^{(2)} + (V_1 - E_i^{(1)}) u_i^{(1)} + (V_2 - E_i^{(2)}) \psi_i^{(0)} = 0.$$

This equation and the following equation for $\psi_i^{(2)}$

$$(H_0 - E_i^{(0)}) \psi_i^{(2)} + (V_1 - E_i^{(1)}) \psi_i^{(1)} + (V_2 - E_i^{(2)}) \psi_i^{(0)} = 0$$

leads to

$$(H_0 - E_i^{(0)}) (u_i^{(2)} - \psi_i^{(2)}) + (V_1 - E_i^{(1)}) (u_i^{(1)} - \psi_i^{(1)}) = 0$$

which, after using Eq. (2.16.1), becomes

$$(H_0 - E_i^{(0)}) (u_i^{(2)} - \psi_i^{(2)}) + (V_1 - E_i^{(1)}) \delta_i^{(1)} \psi_i^{(0)} = 0. \quad (2.17)$$

We then multiply Eq. (2.9.1) by $\delta_i^{(1)}$ and subtract this result from Eq. (2.17); we get

$$(H_0 - E_i^{(0)}) (u_i^{(2)} - \psi_i^{(2)} - \delta_i^{(1)} \psi_i^{(1)}) = 0$$

from which it follows

$$u_i^{(2)} = \psi_i^{(2)} + \delta_i^{(1)} \psi_i^{(1)} + \delta_i^{(2)} \psi_i^{(0)}. \quad (2.16.2)$$

In general, we can easily prove

$$\epsilon_i^{(n)} = E_i^{(n)} \quad (2.15.n)$$

$$u_i^{(n)} = \sum_{j=0}^n \delta_i^{(j)} \psi_i^{(n-j)}. \quad (2.16.n)$$

Thus we see that the orthogonality constraint Eq. (2.10) is really not necessary; it is only a matter of convenience rather than necessity. Adoption of such a constraint only leads to simplification of the eigenfunction corrections. The eigenvalue corrections are not altered in any way by such a man-made constraint which has neither mathematical nor physical basis. This independence of the constraint Eq. (2.10) will be shown in Sect. IV to be related to gauge invariance of the eigenvalue corrections.

III. GAUGE INVARIANCE: WITHOUT MAGNETIC FIELD

In this and the next sections, it will be shown that the perturbation method of the previous section can be made gauge invariant when the operators involved are of the type in Eq. (1.1). The desired gauge invariance can be achieved only if the following identifications are made.³²

$$H_0 = \frac{1}{2m} \hat{p}^2 + eV_0, \quad (3.1)$$

$$V_1 = -\frac{e}{2mc} (\hat{A} \cdot \hat{p} + \hat{p} \cdot \hat{A}), \quad (3.2)$$

$$V_2 = \frac{e^2}{2mc^2} \hat{A} \cdot \hat{A}. \quad (3.3)$$

Here it must be emphasized that the V_2 in Eq. (3.3) remains unchanged even when $\hat{A}$ is not a function of the position vector $\hat{r}$ under some approximation.

In this section, we only investigate the situations when the vector potential is curlless, i.e.,

$$\nabla \times \hat{A} = 0. \quad (3.4)$$

Our justification for working on such a simple case is that we strongly believe that any theory must work first for trivial examples.

A, General Curlless Vector Potential

It will be shown in this subsection that $E_i^{(n)}$ vanishes for all $n \geq 1$ and for all i , regardless of the explicit form of the vector potential as long as it is curlless.

Because of Eq. (3.4), the vector potential can be written as the gradient of a scalar function.

$$A = \nabla \chi(\mathbf{r}, t), \quad (3.5)$$

where the scalar function $\chi(\mathbf{r}, t)$, which will be referred to as a gauge function, must possess all the nice properties so that Eq. (3.4) is satisfied. Using Eq. (3.5), we observe that V_1 and V_2 become greatly simplified.

$$V_1 = \tilde{v}_1, \quad (3.6)$$

$$V_2 = \tilde{v}_2, \quad (3.7)$$

where

$$\tilde{v}_1 \equiv [H_0, \Lambda], \quad (3.8)$$

$$\tilde{v}_2 \equiv \frac{1}{2}[\tilde{v}_1, \Lambda], \quad (3.9)$$

$$\Lambda \equiv -ie \chi(\mathbf{r}, t)/\hbar c. \quad (3.10)$$

In this section, $\tilde{v}_1$ and $\tilde{v}_2$ will take the places of V_1 and V_2 which will be reserved for use in more complicated situations later on. We also mention that the orthogonality constraint will be imposed here.

Now we want to show that

$$E_i^{(n)} = \langle \psi_i^{(0)} | [H_0, \frac{\Lambda^n}{n!}] | \psi_i^{(0)} \rangle = 0 \quad (3.11)$$

for all $n \geq 1$ and for all i . To show this, we must first prove the identity

$$\frac{1}{n!} \Lambda^n \tilde{v}_1 + \frac{1}{(n-1)!} \Lambda^{n-1} \tilde{v}_2 = \frac{1}{(n+1)!} [H_0, \Lambda^{n+1}]. \quad (3.12)$$

The proof of Eq. (3.12) is as follows. The RHS is first expanded

$$\frac{1}{(n+1)!} [H_0, \Lambda^{n+1}] = \frac{1}{(n+1)!} \{ \Lambda^n \tilde{v}_1 + \Lambda^{n-1} \tilde{v}_2 \Lambda + \Lambda^{n-2} \tilde{v}_1 \Lambda^2 + \dots + \Lambda \tilde{v}_1 \Lambda^{n-1} + \tilde{v}_1 \Lambda^n \}. \quad (3.13)$$

We then use the fact that $\tilde{v}_2$ and Λ commute to obtain

$$[\tilde{v}_1, \Lambda^m] = m[\tilde{v}_1, \Lambda]\Lambda^{m-1} = 2m\tilde{v}_2\Lambda^{m-1}. \quad (3.14)$$

Substitution of Eq. (3.14) into (3.13) gives

$$\begin{aligned} \frac{1}{(n+1)!} [H_0, \Lambda^{n+1}] &= \frac{1}{(n+1)!} \{ \Lambda^n \tilde{v}_1 + (\Lambda^{n-1} [\tilde{v}_1, \Lambda] + \Lambda^n \tilde{v}_1) + \dots \} \\ &= \frac{1}{n!} \Lambda^n \tilde{v}_1 + \frac{1}{(n+1)!} \sum_{m=1}^n 2m \Lambda^{n-1} \tilde{v}_2 \\ &= \frac{1}{n!} \Lambda^n \tilde{v}_1 + \frac{1}{(n-1)!} \Lambda^{n-1} \tilde{v}_2. \end{aligned}$$

The last line is the L.H.S. of Eq. (3.12). Thus our proof is completed.

Now we want to prove Eq. (3.11) by induction. For $n = 1$, we get

$$E_i^{(1)} = \langle \psi_i^{(0)} | \tilde{v}_1 | \psi_i^{(0)} \rangle = \langle \psi_i^{(0)} | [H_0, \Lambda] | \psi_i^{(0)} \rangle = 0.$$

Next assume that we have succeeded in showing that

$$E_i^{(1)} = E_i^{(2)} = \dots = E_i^{(m)} = 0, \quad (3.15)$$

then

$$\begin{aligned} E_i^{(m+1)} &= \langle \psi_i^{(0)} | \tilde{v}_1 | \psi_i^{(m)} \rangle + \langle \psi_i^{(0)} | \tilde{v}_2 | \psi_i^{(m-1)} \rangle \\ &= \langle \psi_i^{(0)} | \Lambda(E_i^{(0)} - H_0) | \psi_i^{(m)} \rangle + \langle \psi_i^{(0)} | \tilde{v}_2 | \psi_i^{(m-1)} \rangle \\ &= \langle \psi_i^{(0)} | \Lambda \tilde{v}_1 + \tilde{v}_2 | \psi_i^{(m-1)} \rangle + \langle \psi_i^{(0)} | \Lambda \tilde{v}_2 | \psi_i^{(m-2)} \rangle \\ &= \langle \psi_i^{(0)} | [H_0, \frac{\Lambda^2}{2!}] | \psi_i^{(m-1)} \rangle + \langle \psi_i^{(0)} | \Lambda \tilde{v}_2 | \psi_i^{(m-2)} \rangle. \end{aligned} \quad (3.16)$$

In our arriving at the last step, Eqs. (3.12) and (3.15) have been used. This procedure is then repeated over and over again until

$$E_i^{(m+1)} = \langle \psi_i^{(0)} | [H_0, \frac{\Lambda^{m+1}}{(m+1)!}] | \psi_i^{(0)} \rangle = 0 \quad (3.17)$$

is obtained.

What we have just shown is the following. The identifications of V_1 and V_2 by Eqs. (3.2) and (3.3) enable the RSTIPT to yield identically vanishing perturbation corrections for the eigenvalues when the vector potential is curless.

The question is: Is this result reasonable? It is in the next subsection where we try to answer this question.

B. Causality and Perturbation Corrections

WHY is it so important to have all the perturbation corrections of the eigenenergies vanish identically when the vector potential is curlless?

We wish to answer this question from two different points of view, the mathematical and the physical ones. The explanation to be given here should not be construed to be the only valid reasons. They are only among some important ones. As usual, we start from the mathematical aspect of the problem.

When Eq. (3.4) is satisfied, the vector potential A can be written as the gradient of a "smooth" scalar function as shown in Eq. (3.5). The basis-defining Hamiltonian H_B then has the form

$$H_B = \frac{1}{2m} \left(p - \frac{e}{c} \nabla \chi \right)^2 + eV_0. \quad (3.18)$$

This operator, as can be seen by inspection, is similar³³ to the "unperturbed" Hamiltonian H_0 in Eq. (3.1). Or more explicitly, H_B and H_0 are related by

$$H_B = \exp[ie\chi(\mathbf{r},t)/\hbar c] H_0 \exp[-ie\chi(\mathbf{r},t)/\hbar c]. \quad (3.19)$$

Because of Eq. (3.19), the exact, absolute--as opposed to relative--eigenenergies H_B are identical to those of H_0 , namely $\{E_i^{(0)}\}$. But since we are using H_0 as the zero'th order approximation of H_B , the zero'th order spectrum of H_B therefore is the exact spectrum $\{E_i\}$. This means that

$$E_i = E_i^{(0)} \quad (3.20)$$

and

$$\sum_{n=1}^{\infty} E_i^{(n)} = 0 \quad (3.21)$$

for all i . Because of Eq. (3.20), the only way that each $E_i^{(n)}$, $n \geq 1$, can

represent a "genuine"--as opposed to spurious--perturbation correction is that each vanishes identically, i.e.,

$$E_i^{(n)} = 0, \quad (3.22)$$

for all i and all n ≥ 1 , since mathematically there is no need to modify the zero'th order spectrum which is already exact.

So far we have only "argued passionately" without much compelling reasons because there is no rule prohibiting the use of a bad approximation with spurious and unnecessary corrections. In what follows, we shall present a physical argument which, we hope, will be strong enough to convince the reader that gauge invariance in TIPT is not only a preference but an absolute necessity as well.

In short, we shall show that violation of gauge invariance means violation of causality.³⁴

How to link violation of gauge invariance to violation of causality? We must bring in the propagation characteristics of the potentials in the complete α -Lorentz gauge investigated previously.⁶ Here, we only quote the defining equations and results; the reader is suggested to go to Ref. 6 for a detailed account of the physical properties of electromagnetic gauge transformations.

The complete α -Lorentz gauge is defined by

$$\nabla \cdot \mathbf{A}^{(\alpha L)} + \frac{1}{\alpha c} \frac{\partial}{\partial t} \phi^{(\alpha L)} = 0. \quad (3.23)$$

The most general solutions to Maxwell's equations in this gauge are

$$\begin{aligned} \phi^{(\alpha L)}(\mathbf{r}, t) &= \phi_{\alpha c}^{(\alpha L)}(\mathbf{r}, t) - \frac{1}{c} \frac{\partial}{\partial t} \chi^{(\alpha L)}(\mathbf{r}, t), \\ \mathbf{A}^{(\alpha L)}(\mathbf{r}, t) &= \mathbf{A}_{\alpha c}^{(\alpha L)}(\mathbf{r}, t) + \mathbf{A}_{\alpha c}^{(\alpha L)}(\mathbf{r}, t) + \nabla \chi^{(\alpha L)}(\mathbf{r}, t). \end{aligned} \quad (3.24)$$

All the quantities in Eq. (3.24) are defined below.

$$\phi_{\alpha c}^{(\alpha L)}(\mathbf{x}, t) \equiv \int d\mathbf{x}' \int dt' G(\mathbf{x}, t | \alpha c | \mathbf{x}', t') \rho_{\text{ex}}(\mathbf{x}', t'), \quad (3.25)$$

$$A_{\alpha c}^{(\alpha L)}(\mathbf{x}, t) \equiv -c \int_{t_0}^t dt' \nabla \phi_{\alpha c}^{(\alpha L)}(\mathbf{x}, t'), \quad (3.26)$$

$$A_c^{(\alpha L)}(\mathbf{x}, t) \equiv A_c^{(L)}(\mathbf{x}, t) + c \int_{t_0}^t dt' \nabla \phi_c^{(L)}(\mathbf{x}, t'), \quad (3.27)$$

$$A_c^{(L)}(\mathbf{x}, t) \equiv \frac{1}{c} \int d\mathbf{x}' \int dt' G(\mathbf{x}, t | c | \mathbf{x}', t') J_{\text{ex}}(\mathbf{x}', t'), \quad (3.28)$$

$$\phi_c^{(L)}(\mathbf{x}, t) = [\phi_{\alpha c}^{(\alpha L)}]_{\alpha=1}, \quad (3.29)$$

with

$$G(\mathbf{x}, t | s | \mathbf{x}', t') \equiv \frac{\delta\left(t' + \frac{|\mathbf{x} - \mathbf{x}'|}{s} - t\right)}{|\mathbf{x} - \mathbf{x}'|}, \quad s = c, \alpha c. \quad (3.30)$$

The quantities ρ_{ex} and J_{ex} are the charge and current densities of the external sources which are also presumed to vanish before and at time t_0 . The function $\chi^{(\alpha L)}$ is any function satisfying the differential equation

$$\left(\nabla^2 - \frac{1}{\alpha^2 c^2} \frac{\partial^2}{\partial t^2} \right) \chi^{(\alpha L)} = 0. \quad (3.31)$$

Throughout this paper, $\chi^{(\alpha L)}$ will be chosen to vanish because its presence affects only the initial and boundary conditions which have no physical content.⁶

Here let us explain the meanings of the superscripts and subscripts appearing in Eqs. (3.23) to (3.29). The superscripts (L) and (αL) refer to different values of α in Eq. (3.23); (L) means $\alpha = 1$. The subscripts c and αc indicate that the quantities to which they are attached propagate, respectively, with speed c and αc from the charge and current densities ρ_{ex} and J_{ex} . Since violation of causality is directly tied to how fast physical information propagates through space, these propagation indices therefore are the most relevant quantities.

Now let us investigate what kind of role causality plays in the eigenvalues of the basis-defining Hamiltonians. Assume that we are interested in the results when

$$\alpha > 1.$$

After the sources are turned on, there exists a region which has at least a finite volume such that $A_c^{(L)}$ and $\phi_c^{(L)}$ vanish while $A_{ac}^{(\alpha L)}$ and $\phi_{ac}^{(\alpha L)}$ do not inside this region. This region will be represented symbolically by $S_{c,ac}(t)$. For a charged particle inside this region, the basis-defining Hamiltonian in the α -Lorentz gauge becomes.

$$H_B(A_{ac}^{(\alpha L)}) \equiv \frac{1}{2m} \left(p - \frac{e}{c} A_{ac}^{(\alpha L)} \right)^2 + eV_0 \quad (3.32)$$

where we have slightly altered our notation for future clarity and convenience.

Since the eigenvalues of any basis-defining Hamiltonian are to be interpreted as energy levels, i.e., physical observables, they must not violate causality. In other words, causality demands that $A_{ac}^{(\alpha L)}$, $\alpha > 1$, have no physical effects whatsoever. Mathematically, this means that causality requires that the eigenenergies $\{E_i\}$ of $H_B(A_{ac}^{(\alpha L)})$ be identical to the eigenenergies $\{\epsilon_i\}$ of the operator when $A_{ac}^{(\alpha L)}$ is set to zero, i.e., $H_B(A_{ac}^{(\alpha L)} = 0)$. That is, ³⁵

$$E_i = \epsilon_i \quad \text{for all } i. \quad (3.33)$$

Equation (3.33) is satisfied mathematically since the two operators are similar.³³ We therefore see that the exact spectrum of $H_B(A_{ac}^{(\alpha L)})$ has no conflict with causality.

Now let us turn our attention to the perturbation corrections $\{E_i^{(n)}\}$ of the energy operator H_B in Eq. (3.32). If each correction $E_i^{(n)}$ is to be regarded as being physically meaningful, then causality would certainly require that the presence of $A_{ac}^{(\alpha L)}$, $\alpha > 1$, have absolutely no effects on each perturbation correction, i.e.,

$$E_i^{(n)} = 0 \quad \text{for all } i \text{ and all } n \geq 1. \quad (3.34)$$

It is only after Eq. (3.34) is satisfied can the truncated sum

$$\sum_{n=1}^N E_i^{(n)} \quad (3.35)$$

be physically interpreted, where N is any finite, non-negative integer. Here we must emphasize that the main purpose of any perturbation theory is to use

the truncated sum in Eq. (3.35) as an approximate value for the exact energy. Therefore, in perturbation theory, it is necessary to establish, usually separately, that the approximate values resulted from a certain approximation scheme satisfy the necessary conditions of an observable, among which is causality. We have also learned that Eqs. (3.1), (3.2), and (3.3) enable the RSTIPT to be gauge invariant and to conform to causality when the vector potential in the basis-definition Hamiltonian is curlless.

C. A Special Example

What is so special when the vector potential is not a function of the position variable,

$$A(\mathbf{r}, t) = A(0, t), \quad (3.36)$$

is that this is where all controversies on gauge transformations in quantum mechanics occur, time-independent²⁴⁻²⁷ or time-dependent¹⁵⁻²² cases. As stated before, we concentrate our efforts only on time-independent problems. We are going to prove

$$\Delta_i^{(2)} = - \frac{e^2}{2mc^2} A(0, t) \cdot A(0, t). \quad (3.37)$$

This quantity $\Delta_i^{(2)}$ is given by

$$\Delta_i^{(2)} = \sum_{j \neq i} \frac{\langle \psi_i^{(0)} | V | \psi_j^{(0)} \rangle \langle \psi_j^{(0)} | V | \psi_i^{(0)} \rangle}{E_i^{(0)} - E_j^{(0)}}, \quad (3.38)$$

where

$$V = - \frac{e}{mc} A(0, t) \cdot \mathbf{p}. \quad (3.39)$$

Let us now observe that V can also be written as

$$V = [H_0, \Lambda], \quad (3.40)$$

where Λ now assumes the form

$$\Lambda = -ie \mathbf{r} \cdot A(0, t) / c\hbar. \quad (3.41)$$

Substitution of Eq. (3.40) into Eq. (3.38) leads to

$$\begin{aligned}\Delta_i^{(2)} &= \sum_{j \neq i} \frac{\langle \psi_i^{(0)} | [H_0, \Lambda] | \psi_j^{(0)} \rangle \langle \psi_j^{(0)} | V | \psi_i^{(0)} \rangle}{E_i^{(0)} - E_j^{(0)}} \\ &= \sum_{j \neq i} \langle \psi_i^{(0)} | \Lambda | \psi_j^{(0)} \rangle \langle \psi_j^{(0)} | [H_0, \Lambda] | \psi_i^{(0)} \rangle\end{aligned}\quad (3.42)$$

$$= \sum_j \langle \psi_i^{(0)} | \Lambda | \psi_j^{(0)} \rangle \langle \psi_j^{(0)} | [H_0, \Lambda] | \psi_i^{(0)} \rangle \quad (3.43)$$

$$= \langle \psi_i^{(0)} | \Lambda [H_0, \Lambda] | \psi_i^{(0)} \rangle, \quad (3.44)$$

Here let us call our attention to the summations in Eqs. (3.42) and (3.43). The summation in the former runs through all j except when $j = i$, while the summation of Eq. (3.43) runs through all j including $j = i$. This is because

$$\langle \psi_i^{(0)} | [H_0, \Lambda] | \psi_i^{(0)} \rangle = 0. \quad (3.45)$$

Similarly, we get

$$\Delta_i^{(2)} = - \langle \psi_i^{(0)} | [H_0, \Lambda] \Lambda | \psi_i^{(0)} \rangle. \quad (3.46)$$

Using Eqs. (3.44) and (3.46), we prove Eq. (3.37) as follows.

$$\begin{aligned}\Delta_i^{(2)} &= \frac{1}{2} \Delta_i^{(2)} + \frac{1}{2} \Delta_i^{(2)} \\ &= \frac{1}{2} \langle \psi_i^{(0)} | \Lambda [H_0, \Lambda] - [H_0, \Lambda] \Lambda | \psi_i^{(0)} \rangle \\ &= - \frac{1}{2} \langle \psi_i^{(0)} | [V, \Lambda] | \psi_i^{(0)} \rangle \\ &= - \frac{1}{2} \langle \psi_i^{(0)} | \left[-\frac{e}{mc} \hat{A}(0, t) \cdot \hat{p}, -\frac{ie}{\hbar c} \hat{x} \cdot \hat{A}(0, t) \right] | \psi_i^{(0)} \rangle \\ &= - \frac{e^2}{2mc^2} \hat{A}(0, t) \cdot \hat{A}(0, t).\end{aligned}\quad (3.47)$$

Thus, we see that neglecting the mathematical and physical significance of the $\hat{A} \cdot \hat{A}$ term is not warranted even when this term is a constant. We notice here that the above result is consistent with Eq. (3.11).

IV. GAUGE INVARIANCE: WITH MAGNETIC FIELD

A. General Vector Potential

In the previous section, it has been shown that Eqs. (3.1) to (3.3) enables the RSTIPT to yield identically vanishing perturbation corrections in eigen-energies when the vector potential does represent a vanishing magnetic field. Such results are gauge invariant because they do not depend upon the explicit form of the vector potential as long as it is curless. We have also learned that because of the gauge invariance of the energy corrections, no causality is violated. Although curless potentials are theoretically very important because they serve as a first test case for any theory, they do not represent much interesting physics.

In this section, we shall show that even when

$$\nabla \times \mathbf{A}(\mathbf{r}, t) = \mathbf{B}(\mathbf{r}, t) \neq 0, \quad (4.1)$$

the perturbation corrections given in Eq. (2.9) will still be gauge invariant. It will also be shown that the independence of the eigenvalue corrections from the orthogonality constraint Eq. (2.10) is related to the gauge invariance of these corrections.

Let us first assume that we have two non-identical vector potentials representing the same magnetic field, i.e.,

$$\nabla \times \mathbf{A}(\mathbf{r}, t) = \nabla \times \mathbf{A}'(\mathbf{r}, t) = \mathbf{B}(\mathbf{r}, t); \quad \mathbf{A}(\mathbf{r}, t) \neq \mathbf{A}'(\mathbf{r}, t). \quad (4.2)$$

From elementary knowledge of classical electrodynamics, $\mathbf{A}$ and $\mathbf{A}'$ are related by a gauge function $\chi(\mathbf{r}, t)$ in the following way.

$$\mathbf{A}'(\mathbf{r}, t) = \mathbf{A}(\mathbf{r}, t) + \nabla \chi(\mathbf{r}, t). \quad (4.3)$$

The basis-defining Hamiltonians in these two gauges are, respectively,¹⁹

$$\begin{aligned} H_B &= \frac{1}{2m} \left(\mathbf{p} - \frac{e}{c} \mathbf{A} \right)^2 + eV_0, \\ H'_B &= \frac{1}{2m} \left(\mathbf{p} - \frac{e}{c} \mathbf{A}' \right)^2 + eV_0. \end{aligned} \quad (4.4)$$

According to Eqs. (3.1) to (3.3), H_B and H'_B are decomposed into three pieces.³⁶

$$H_B = H_0 + V_1 + V_2, \quad (4.5)$$

$$H'_B = H_0 + V'_1 + V'_2,$$

where V_1 and V_2 are defined in Eqs. (3.2) and (3.3), and

$$V'_1 = -\frac{e}{2mc}(\mathbf{A}' \cdot \mathbf{p} + \mathbf{p} \cdot \mathbf{A}') \quad (4.6)$$

$$V'_2 = \frac{e^2}{2mc^2} \mathbf{A}' \cdot \mathbf{A}'. \quad (4.7)$$

Using Eqs. (3.8), (3.9), (3.10), and (4.3), we obtain

$$V'_1 = V_1 + \tilde{V}_1 \quad (4.8)$$

$$V'_2 = V_2 + \tilde{V}_2 + \hat{V}_2 \quad (4.9)$$

where

$$\hat{V}_2 = [V_1, \Lambda]. \quad (4.10)$$

Let us now denote by $\{E_i^{(n)}\}$ and $\{\psi_i^{(n)}\}$ the eigenvalues and corrections and the corresponding eigenfunctions and corrections of H_B , and by $\{\epsilon_i^{(n)}\}$ and $\{\phi_i^{(n)}\}$ those of H'_B . $\{E_i^{(n)}\}$ and $\{\psi_i^{(n)}\}$ satisfy Eqs. (2.9) and (2.11) with H_0 , V_1 and V_2 given by Eqs. (3.1) to (3.3). For convenience, the orthogonality constraint Eq. (2.10) is imposed upon both $\{\psi_i^{(n)}\}$ and $\{\phi_i^{(n)}\}$. $\{\epsilon_i^{(n)}\}$ and $\{\phi_i^{(n)}\}$ are to be obtained from the following equations.

$$H_0 \phi_i^{(0)} = \epsilon_i^{(0)} \phi_i^{(0)} \quad (4.11.0)$$

$$H_0 \phi_i^{(1)} + V'_1 \phi_i^{(0)} = \epsilon_i^{(0)} \phi_i^{(1)} + \epsilon_i^{(1)} \phi_i^{(0)} \quad (4.11.1)$$

$$H_0 \phi_i^{(n)} + V'_1 \phi_i^{(n-1)} + V'_2 \phi_i^{(n-2)} = \sum_{j=0}^n \epsilon_i^{(j)} \phi_i^{(n-j)} \quad (4.11.n)$$

and

$$\epsilon_i^{(1)} = \langle \phi_i^{(0)} | V'_1 | \phi_i^{(0)} \rangle \quad (4.12.1)$$

$$\epsilon_i^{(n)} = \langle \phi_i^{(0)} | V'_1 | \phi_i^{(n-1)} \rangle + \langle \phi_i^{(0)} | V'_2 | \phi_i^{(n-2)} \rangle - \sum_{j=1}^{n-1} \epsilon_i^{(j)} \langle \phi_i^{(0)} | \phi_i^{(n-j)} \rangle. \quad (4.12.n)$$

Here we must call to the attention that Eqs. (4.11) and (4.12) are identical in form to Eqs. (2.9) and (2.11), respectively. That is, when the substitutions

$$\epsilon_i^{(n)} \rightarrow E_i^{(n)}, \quad \phi_i^{(n)} \rightarrow \psi_i^{(n)}, \quad \Lambda_i', V_1', V_2' \rightarrow \Lambda_i, V_1, V_2$$

are made, Eqs. (4.11) and (4.12) are indistinguishable from Eqs. (2.9) and (2.11). This is what we meant in Introduction by "identical procedures."

Our goal is to prove that

$$\epsilon_i^{(n)} = E_i^{(n)} \quad \text{for all } i \text{ and all } n. \quad (4.13)$$

From Eqs. (2.9.0) and (4.11.0), it is clear that Eq. (4.13) is true for $n = 0$. It is also clear that

$$\phi_i^{(0)} = \psi_i^{(0)}. \quad (4.14.0)$$

To prove Eq. (4.13) systematically, we first introduce the functions $\{u_i^{(n)}\}$ by

$$u_i^{(n)} \equiv \sum_{j=0}^n \frac{\Lambda_i^j}{j!} \phi_i^{(n-j)}, \quad n \geq 0. \quad (4.15)$$

These functions, as will be shown later, satisfy the following equations.

$$H_0 u_i^{(0)} = \epsilon_i^{(0)} u_i^{(0)} \quad (4.16.0)$$

$$H_0 u_i^{(1)} + V_1 u_i^{(0)} = \epsilon_i^{(0)} u_i^{(1)} + \epsilon_i^{(1)} u_i^{(0)} \quad (4.16.1)$$

$$H_0 u_i^{(n)} + V_1 u_i^{(n-1)} + V_2 u_i^{(n-2)} = \sum_{j=0}^n \epsilon_i^{(j)} u_i^{(n-j)}, \quad (4.16.n)$$

From these equations, we also get

$$\epsilon_i^{(1)} = \langle u_i^{(0)} | V_1 | u_i^{(0)} \rangle \quad (4.17.1)$$

$$\epsilon_i^{(n)} = \langle u_i^{(0)} | V_1 | u_i^{(n-1)} \rangle + \langle u_i^{(0)} | V_2 | u_i^{(n-2)} \rangle - \sum_{j=1}^{n-1} \epsilon_i^{(j)} \langle u_i^{(0)} | u_i^{(n-j)} \rangle. \quad (4.17.n)$$

In fact, the above expressions for $\{\epsilon_i^{(n)}\}$ can also be directly obtained from Eq. (4.12) by repeatedly using Eqs. (4.8) to (4.10).

Now we only have to go through the same procedure outlined in Sect. II.B to see that Eq. (4.13) is true for all n . This is because the only difference between $\{\psi_i^{(n)}\}$ and $\{u_i^{(n)}\}$ are the orthogonality constraint Eq. (2.10) which

has been shown to be irrelevant to the eigenvalue corrections. The only thing that need be proven is that $\{u_i^{(n)}\}$ satisfy Eq. (4.16). This will be done next.

B. Derivation of Eq. (4.16)

In order that Eq. (4.13) be true, Eq. (4.16) must be derived. There is no question that Eq. (4.16.0) is true. Let us now consider $n = 1$. Substituting the definitions of $u_i^{(0)}$ and $u_i^{(1)}$ into the LHS of Eq. (4.16), we get

$$\begin{aligned} H_0 u_i^{(1)} + v_1 u_i^{(0)} &= H_0 (\phi_i^{(1)} + \Lambda \phi_i^{(0)}) + v_1 \phi_i^{(0)} \\ &= H_0 \phi_i^{(0)} + [H_0, \Lambda] \phi_i^{(0)} + \Lambda H_0 \phi_i^{(0)} + v_1 \phi_i^{(0)} \\ &= H_0 \phi_i^{(1)} + (\tilde{v}_1 + v_1) \phi_i^{(0)} + \Lambda \epsilon_i^{(0)} \phi_i^{(0)}. \end{aligned}$$

We then make use of Eq. (4.11.1) to get

$$\begin{aligned} H_0 u_i^{(1)} + v_1 u_i^{(0)} &= \epsilon_i^{(0)} \phi_i^{(1)} + \epsilon_i^{(0)} \Lambda \phi_i^{(0)} + \epsilon_i^{(1)} \phi_i^{(0)} \\ &= \epsilon_i^{(0)} u_i^{(1)} + \epsilon_i^{(1)} u_i^{(0)}. \end{aligned}$$

Thus we have shown Eq. (4.16.1).

How to obtain Eq. (4.16.n) is as follows. We first consider the first term.

$$\begin{aligned} H_0 u_i^{(n)} &= \sum_{j=0}^n \frac{\Lambda^j}{j!} H_0 \phi_i^{(n-j)} + \sum_{j=0}^n \left[H_0, \frac{\Lambda^j}{j!} \right] \phi_i^{(n-j)} \\ &= - \sum_{j=0}^{n-1} \frac{\Lambda^j}{j!} v_1' \phi_i^{(n-1-j)} - \sum_{j=0}^{n-2} \frac{\Lambda^j}{j!} v_2' \phi_i^{(n-2-j)} + \sum_{j=0}^n \sum_{k=0}^{n-j} \frac{\Lambda^j}{j!} \phi_i^{(n-j-k)} \epsilon_i^{(k)} \\ &\quad + \sum_{j=1}^n \frac{\Lambda^{j-1}}{(j-1)!} \tilde{v}_1 \phi_i^{(n-j)} + \sum_{j=2}^n \frac{\Lambda^{j-2}}{(j-2)!} \tilde{v}_2 \phi_i^{(n-j)} \end{aligned} \quad (4.18)$$

where Eqs. (4.11.n) and (3.12) have been used. Changing the summation limits of the last three terms yields

$$\begin{aligned}
H_0 u_i^{(n)} &= - \sum_{j=0}^{n-1} \frac{\Lambda^j}{j!} (v_1' - \tilde{v}_1) \phi_i^{(n-1-j)} - \sum_{j=0}^{n-2} \frac{\Lambda^j}{j!} (v_2' - \tilde{v}_2) \phi_i^{(n-2-j)} \\
&\quad + \sum_{k=0}^n \sum_{j=0}^{n-k} \epsilon_i^{(k)} \frac{\Lambda^j}{j!} \phi_i^{(n-k-j)} \\
&= - \sum_{j=0}^{n-1} \frac{\Lambda^j}{j!} v_1 \phi_i^{(n-1-j)} - \sum_{j=0}^{n-2} \frac{\Lambda^j}{j!} (v_2 + \tilde{v}_2) \phi_i^{(n-2-j)} + \sum_{k=0}^n \epsilon_i^{(k)} u_i^{(n-k)},
\end{aligned} \tag{4.19}$$

where we have used Eqs. (4.8), (4.9) and (4.15). After regrouping, we get

$$\begin{aligned}
H_0 u_i^{(n)} &= -v_1 \sum_{j=0}^{n-1} \frac{\Lambda^j}{j!} \phi_i^{(n-1-j)} + \sum_{j=0}^{n-1} \left[v_1, \frac{\Lambda^j}{j!} \right] \phi_i^{(n-1-j)} \\
&\quad - v_2 \sum_{j=0}^{n-2} \frac{\Lambda^j}{j!} \phi_i^{(n-2-j)} - \sum_{j=0}^{n-2} \tilde{v}_2 \frac{\Lambda^j}{j!} \phi_i^{(n-2-j)} + \sum_{k=0}^n \epsilon_i^{(k)} u_i^{(n-k)} \\
&= -v_1 u_i^{(n-1)} - v_2 u_i^{(n-2)} + \sum_{j=0}^n \epsilon_i^{(j)} u_i^{(n-j)},
\end{aligned}$$

since

$$\left[v_1, \frac{\Lambda^j}{j!} \right] = \frac{\Lambda^{j-1}}{(j-1)!} \tilde{v}_2. \tag{4.20}$$

Hence, our proof of Eq. (4.13) is completed.

C. Causality and Gauge Invariance

Let us once again investigate the relationship between gauge invariance and causality. As usual, we must bring into discussions the propagation characteristics of classical electromagnetic gauge transformations. Our investigation can be best elucidated by considering

$$A_\alpha(\underline{x}, t) = A(\underline{x}, t) + A_{\alpha c}^{(uL)}, \quad \alpha > 1. \tag{4.21}$$

Let us point out once again that $A_{\alpha c}^{(uL)}$ propagates with speed $\alpha c > c$ from the charge and current densities. This term must not cause any change in physical quantities since if otherwise it would imply that there is some way that our

quantum system can "receive message" that propagates through space with speed faster than c , the speed of light in vacuum, from the source of this message.

As can be obviously seen by inspection, the basis-defining Hamiltonians H_B and H'_B are similar since this causality violating potential is curless. Mathematically, the exact eigenenergies of these two operators are therefore identical. Hence, as far as the exact physics is concerned, nothing need be done because operator algebra already ensures that no causality is violated.

This is not true, however, with perturbation theories. By design, a perturbation theory is to solve an equation, which in our situation is an eigenvalue equation, by successive approximations. Furthermore, the steps of approximations always terminate at finite numbers. Associated with each finite step of approximation is an approximate value for the quantity in question. The problem becomes whether the approximate value resulting from any finite step of approximations can be physically interpreted. In order that it can be, it must satisfy all necessary conditions of an observable for any finite terms of the sum

$$\sum_{n=0}^N \epsilon_i^{(n)}$$

where N is any non-negative integer. From this, it is therefore inferred that each perturbation correction must satisfy all the necessary conditions of an observable. Causality is only one necessary condition that any physical observable must not violate.

We have seen that the RSTIPT is gauge invariant. This means any gradient component in the vector potential causes no change in the eigenenergy corrections. Hence, the vector potential $A_{\alpha c}^{(\alpha L)}$, $\alpha \neq 1$, does not have any consequences in the eigenvalue correction. It is then immediately clear that gauge invariance ensures that causality is not violated by each order of these corrections.

V. CONCLUSIONS

We have shown that, if properly classifying the perturbation interactions, the Rayleigh-Schrödinger time-independent perturbation theory is gauge invariant when the operators concerned are the basis-defining Hamiltonians introduced previously. We have also proven that the orthogonality constraint that is usually imposed upon the eigenfunction corrections is really not necessary although its imposition leads to considerable simplification for the eigenfunction corrections. The independence of the eigenvalue corrections from the orthogonality constraint is a must since there is no compelling mathematical and physical reasons to impose such an arbitrary constraint on quantities through which physical observables are to be calculated. We have also established that the fact that the eigenvalue corrections to all orders are independent of the projections of the eigenfunction corrections into their own zero'th order eigenfunctions is directly related to the fact that these eigenvalue corrections are independent of the choice of gauge.

One important reason for demanding gauge invariance for all physical observables is that violation of gauge invariance can be related directly to violation of causality.³⁴ Although gauge invariance has been a subject of investigation for several decades, it is possible only very recently to link gauge invariance to causality after the discovery that there is a characteristic propagation behavior associated with each particular gauge.⁴⁻⁶ The introduction of the complete α -Lorentz gauge and the complete, closed form solutions of potentials in this gauge⁶ make it even more obvious why there is no excuse not to demand gauge invariance for all physical observables.

This work was supported in part by the National Science Foundation and the U. S. Energy Research and Development Administration. The author wishes to thank the Nuclear Theory Group of the University of Maryland for the hospitality and encouragement during the past two years.

REFERENCES

[†]Present address: Theoretical Chemistry Institute, University of Wisconsin, Madison, Wisconsin, 53706.

1. Only by defining gauges (of the second kind) in this way can there be a one-to-one correspondence between gauge transformations of the first kind and those of the second kind.
2. The most famous and well-known examples are on the Coulomb gauge formulation of quantum electrodynamics by E. Fermi, Rev. Mod. Phys. 4, 87 (1932), and J. Schwinger, Phys. Rev. 74, 1439 (1948). Other examples abound throughout literature and textbooks.
3. The reader should be aware that recent progress in the understanding of the physical properties of electromagnetic gauges, notably on the propagation behavior of potentials, may not be completely compatible with the traditional understanding (for example, Ref. 2) on the subject of classical gauge transformations. For more information, see Refs. 4 to 6.
4. O. L. Brill and B. Goodman, Amer. J. Phys. 35, 832 (1967).
5. J. D. Jackson, Classical Electrodynamics (Wiley, New York, 1975, 2nd ed.), pp. 220-223.
6. K.-H. Yang, Ann. Phys. (N.Y.) 101, 97 (1976).
7. W. E. Lamb, Jr., Phys. Rev. 85, 259 (1952).
8. B. S. DeWitt, Phys. Rev. 125, 2189 (1962).
9. S. Mandelstam, Ann. Phys. 19, 1 (1962).
10. F. J. Belinfante, Phys. Rev. 128, 2832 (1962).
11. F. Rohrlich and F. Strocchi, Phys. Rev. 139, B476 (1965).
12. R. J. Torrence and M. M. Tulczyjew, J. Math. Phys. 14, 1725 (1973).
13. F. Strocchi and A. S. Wightman, J. Math. Phys. 15, 2198 (1974).

14. R. Menikoff and D. H. Sharp, J. Math. Phys. 18, 471 (1977).
15. For papers colliding head-on on the subject (or if we prefer, controversy) of $A(Q,t) \cdot p$ vs. $E(Q,t) \cdot r$, first raised by Lamb (Ref. 7) in connection with experimental measurement, see Ref. 7 and Refs. 16 to 22.
16. E. A. Power and S. Zienan, Phil. Trans. Roy. Soc. A251, 427 (1959).
17. Z. Fried, Phys. Rev. A8, 2835 (1973).
18. R. G. Woolley, Adv. Chem. Phys. 33, 153 (1975).
19. K.-H. Yang, Ann. Phys. (N.Y.) 101, 62 (1976).
20. K.-H. Yang, Univ. of Maryland Tech. Rpt. #77-019 (1976).
21. D. H. Kobe and A. L. Smirl, preprint (1977).
22. E. A. Power and T. Thirunamachandran, Phys. Rev. A15, 2366 (1977).
23. We only confine ourselves to situations where electromagnetic fields and potentials are treated as classical and external quantities. Throughout this paper, the terms "gauge" and "Quantum Mechanics" are used only in this context. For a new approach to the gauge problems in quantum mechanics see Refs. 6, 19 and 20.
24. Recently, the trick of transforming from one gauge to another has been used again for computing the energy shift of an atomic system in the presence of a strong, external electromagnetic field. See, for more information and controversy, for example, Refs. 25 to 27.
25. R. F. O'Connell, Phys. Rev. A12, 1132 (1975).
26. G. W. Ford and R. F. O'Connell, Phys. Rev. A13, 1281 (1976).
27. P. Lambropoulos, E. A. Power, and T. Thirunamachandran, Phys. Rev. A14, 1910 (1976).
28. Whether or not an approximation method is gauge invariant has started being recognized to be very important, for example, in the atomic structure theory. See, for example, D. L. Lin, Phys. Rev. A16, 600 (1977).

29. Here, we stress that one should distinguish the subtle and often confusing differences among the following terms in the discussions of "the gauge problems" in quantum mechanics. They are: (1) gauge dependence, (2) gauge invariance, (3) gauge independence, (4) explicit gauge dependence, and (5) explicit gauge independence. Unlike Strocchi and Wightman (Ref. 13) who give formal definitions to the terms they use, we only give examples to illustrate what we mean here. Note, all of the examples here are valid only in the Schrödinger representation. To use $-i\hbar\nabla$ for the canonical momentum is gauge independent (see E. Nelson, Phys. Rev. 150, 1079 (1966)) because it has nothing to do with the gauge in the Hamiltonian. Therefore $p(\equiv -i\hbar\nabla)$ is an explicitly gauge independent operator. But it is gauge dependent, or if one prefers, implicitly gauge dependent, because its expectation value does depend on the choice of gauge (Refs. 19, 20). (In other language, it is not form invariant under gauge transformations of the first kind; see Ref. 21). Another operator of this category is the "unperturbed" Hamiltonian $H_0 = \frac{1}{2m} p^2 + eV_0$. The operator $(m\dot{x}) \equiv p - \frac{e}{c} A$ ($\equiv -i\hbar\nabla - \frac{e}{c} A$) is explicitly gauge dependent because p is explicitly gauge independent but A is by definition gauge dependent. But this operator is gauge invariant because its expectation value is gauge invariant (Ref. 19, and see, for example, D. Bohm, Quantum Theory (Prentice-Hall, New York, 1951), pp. 357-359). (Other statement is $(m\dot{x})$ as defined above is form invariant under gauge transformations, see Ref. 21.) Another operator of this category is the basis-defining Hamiltonian $H_B \equiv \frac{1}{2m} (m\dot{x})^2 + eV_0$.

In classical mechanics or in Heisenberg's representation, a canonical momentum is not explicit gauge independent. In fact, it is gauge dependent, the reason being $p = m\dot{x} + \frac{e}{c} A$ (for a nonrelativistic, spinless charged particle). The vector potential A is gauge dependent by definition while $m\dot{x}$

is gauge invariant. The only conclusion is p is gauge dependent. Another way to realize the gauge dependence of a canonical momentum is by $p_H = U^+ (\frac{\hbar}{i} \nabla) U$ where U is a time-evolution operator which is the solution of the differential equation $i\hbar \frac{\partial}{\partial t} U = HU$. Since H is gauge dependent (otherwise the Schrödinger equation would not be gauge invariant, also see Refs. 19, 21), U is gauge dependent. Since $-i\hbar \nabla$ is explicitly gauge independent p_H is therefore gauge dependent. The operator $(m\gamma)_H$ is still gauge invariant.

30. See, for example, (a) A. Messiah, Quantum Mechanics (North-Holland Co., Amsterdam, 1966), Chapt. 16, and (b) E. Merzbacher, Quantum Mechanics (Wiley, New York, 1961), Chapt. 16.
31. H stands for a Hermitian operator, not for a Hamiltonian.
32. We stress that Eqs. (3.1), (3.2) and (3.3) are not new discoveries. These can be easily realized when A is quantized.
33. We use the conventional language that two operators A and B are similar if $A = SBS^{-1}$ where S is a nonsingular transformation.
34. Direct relation between violation of causality and violation of gauge invariance in quantum mechanical interpretation of probability amplitudes is first observed and established in Ref. 20. That the Coulomb gauge has a causality violating component in the vector potential is first observed by Brill and Goodman in 1967 (Ref. 4) and independently by the author in 1976 (Ref. 6).
35. A more general requirement from causality is that the spectrum of $H_B(A + A_{vac}^{(\alpha L)})$ be identical to that of $H_B(A)$ for any arbitrary A as long as $\alpha \neq 1$ (in vacuum only). Similarly for the eigenvalue corrections of all orders.
36. We emphasize here in Eqs. (4.4) and (4.5) that H'_B , V'_1 and V'_2 MUST be obtained by replacing A by A' in H_B , V_1 , and V_2 , respectively.