LIMITATIONS OF GAUGE INVARIANCE
and
CONSEQUENCES FOR LASER-INDUCED PROCESSES

This lecture is based on a talk given at the
Joint Theoretical Colloquium of ETH Zurich and University of Zürich on 17 May 2010.

• Physical processes involving electromagnetic interactions depend entirely on the electric and magnetic fields that are present.

• These fields can be represented as derivatives of scalar and vector potential functions that are often mathematically convenient, but are actually only auxiliary quantities. (Each such set is called a *gauge*.)

• The potential functions are not unique. There are many sets of these functions corresponding to the same fields and thus to the same physical processes. (They are connected by *gauge transformations.)*

• The choice of which of the possible sets of potentials to use for a given problem is a matter of mathematical or conceptual convenience.

• Exactly the same considerations about the choice of gauge apply to both classical and quantum phenomena.

Not one of the above statements is entirely correct.

Furthermore…
There is another set of statements that is accepted almost universally in the Atomic, Molecular, and Optical (AMO) physics community, and to some extent outside that community:

• A gauge transformation is a unitary transformation.
• Quantum transition amplitudes are manifestly gauge invariant.

Less universal, but also widespread:
• The most basic approach is to formulate all problems using a scalar potential. (Within the dipole approximation, this is called the *length gauge.*) If a different gauge is desired, this should be obtained through a gauge transformation from the length gauge.

All of the bulleted statements above are false.

- Review the rules for gauges and gauge transformations.
- Examine an elementary one-dimensional classical example to illustrate some of the problems to be explored.
- Show how conservation conditions can be changed by gauge transformations.
- Examine the essential differences between Newtonian mechanics and other formulations.
- Examples from strong-field physics. (Many gauge transformation difficulties are exacerbated by strong fields.)
- Exhibit an essential difference between the properties of classical and quantum gauge transformations.
- Show how the above demonstration solves some essential long-standing dilemmas in the foundations of strong-field physics.
- The above considerations infer the existence of a *physical gauge.* (This is still controversial and not essential.)
- Revisit the statements from the Preamble.

ELECTROMAGNETIC POTENTIALS

Electric fields **E** and magnetic fields **B** can be written in terms of scalar potentials ϕ vector potentials **A** through the connections *(Gaussian units)*

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

$$\mathbf{B} = \nabla \times \mathbf{A}$$

When substituted into Maxwell's equations, and with the *Lorentz condition*

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

enforced in order to decouple the ϕ and **A** equations, the result is the wave equations

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\phi = -4\pi\rho$$

$$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{A} = -\frac{4\pi}{c}\mathbf{J}$$

where ρ is a charge density and **J** is a current density.

One of the senior participants in the Joint Theoretical Physics Colloquium expressed a dislike for the Lorentz condition.

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

I disagree, based on its necessity for decoupling the inhomogeneous wave equations for scalar and vector potentials. Arguments based on approximations that relieve the necessity for strict adherence to the Lorentz condition become questionable when fields are strong.

There is a superficial discrepancy in the LG, where $\mathbf{A}$ = 0 and $\partial\phi/\partial t \neq 0$, but this is not essential because there is no wave equation for $\mathbf{A}$ from which to decouple. However, the fact that $\partial\phi/\partial t \neq 0$ **IS** important because the wave equation for ϕ then requires a source density ρ, whereas a plane wave does not require any sources. This just emphasizes the unphysical nature of the LG for description of PW phenomena.

There is a separate issue with the Coulomb gauge. It is stated in many textbooks (including Jackson) that a defining condition for the Coulomb gauge is that $\nabla \cdot A = 0$.
That condition, plus the constraint $\phi = 0$, leads to the Lorentz condition being satisfied.
However, there is no covariant (completely relativistic) equivalent for this, whereas the Lorentz condition in its full form does have a covariant statement.

If χ is any scalar function that satisfies the homogeneous wave equation

$$\nabla^2 \chi - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\chi = 0$$

then the **gauge transformation**

$$\phi \rightarrow \phi' = \phi - \frac{1}{c}\frac{\partial}{\partial t}\chi,$$

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla\chi.$$

leaves the electric and magnetic fields unchanged.

LORENTZ FORCE

A particle of charge *q*, subjected to an electric field **E** and a magnetic field **B** while moving with velocity **v** will experience a force **F**[Lorentz] :

$$\mathbf{F}^{Lorentz} = q\left[\mathbf{E} + \frac{1}{c}(\mathbf{v} \times \mathbf{B})\right]$$

The trajectory of motion of a particle of charge *q* and mass *m* subjected to an electromagnetic field is given by the solution of the Newtonian equation of motion

$$\frac{d}{dt}\mathbf{P} = \mathbf{F}^{Lorentz} = q\left[\mathbf{E} + \frac{1}{c}\left(\mathbf{v} \times \mathbf{B}\right)\right]$$

✓The trajectory depends only on the force.
✓The force depends only on the fields **E** and **B**.
✓The fields are unchanged by a gauge transformation.
✓Therefore the solution of the problem is **gauge invariant**.

The analysis above has long been accepted as conclusive evidence for gauge invariance as a fundamental and unambiguous principle of electrodynamics.

For example:

• The equations of quantum mechanics are all in terms of potentials and not fields. Many attempts have been made to write the Schrödinger equation directly in terms of E and B, but none have succeeded. All such attempts have yielded nonlocal formulations.

• In textbooks on classical mechanics it is common to find demonstrations that alternative formulations (Hamiltonian, Lagrangian, …, expressed in terms of potentials) can be shown to infer Newton's equations, but not the converse.

• In strong-field physics, two gauges are in common use, and they can be shown to contradict each other in fundamental ways.

An elementary example illustrates basic problems with gauge invariance

A CHARGED PARTICLE IN A CONSTANT ELECTRIC FIELD
ONE-DIMENSIONAL TREATMENT

Classical electrostatics is a complete field in itself. It can be done exclusively with scalar potentials.

Elementary example: a particle of mass m and charge q in a constant electric field E_0 . This can be done in one dimension x.

Hamiltonian: $H = p^2/2m - qE_0x$

Hamilton's equations:

$$\dot{x} = \frac{\partial H}{\partial p} = \frac{p}{m} \; ; \quad \dot{p} = -\frac{\partial H}{\partial x} = qE_0$$

These equations can be combined to give:
$$m\ddot{x} = qE_0,$$

This is Newton's equation: (mass)×(acceleration) = force

The solution is elementary; it is quadratic in time.

Now a simple gauge transformation will be applied.
What has been done above corresponds to what is often called the _length gauge_, because the interaction term in the Hamiltonian has the form $-E_0x$. The gauge transformation to be applied eliminates the scalar potential and introduces a one-dimensional version of a vector potential. This is called the _velocity gauge_.

The gauge transformation has the generating function
$$\chi = -cE_0x\,t$$

GAUGE TRANSFORMATION

$$\phi = -xE_0 \quad \rightarrow \quad \phi' = \phi - \frac{1}{c}\frac{\partial \chi}{\partial t} = 0$$

$$A = 0 \quad \rightarrow \quad A' = A + \frac{\partial \chi}{\partial x} = -cE_0 t$$

$$H = \frac{p^2}{2m} - qxE_0 \quad \rightarrow \quad H' = \frac{1}{2m}\left(p' + qE_0 t\right)^2$$

Transformed Hamilton's equations are:

$$\dot{x} = \frac{\partial H'}{\partial p'} = \frac{p' + qE_0 t}{m}, \quad \dot{p}' = -\frac{\partial H'}{\partial x} = 0$$

These equations can be combined to give: $m\ddot{x} = qE_0$

THE EQUATION OF MOTION IS AS BEFORE.

GAUGE INVARIANCE IS SATISFIED!

GAUGE EQUIVALENT – BUT DIFFERENT

The most obvious difference is in the conservation of energy.
Written as *H = kinetic energy + potential energy = T + V*:

$$H = T + V = constant$$

$$H' = T'(t); \quad V' = 0$$

Evaluated with respective solutions: *T = T'* as expected.

This must be so because kinetic energy is measurable in the laboratory and must be preserved by a gauge transformation.

HOWEVER: $V \neq V'$. THIS IS A FUNDAMENTAL DIFFERENCE.

In the original gauge, potential energy is converted into kinetic energy. After the gauge transformation, kinetic energy grows as t^2 with no source for this unlimited energy.

If the Hamiltonian (or Lagrangian) is independent of a particular generalized coordinate, then the conjugate generalized momentum is conserved. This applies in as many generalized coordinates as exist in the problem.

$$\dot{p} = -\frac{\partial H}{\partial x} = 0 \quad \Rightarrow \quad p = \text{const.}$$

In particular, if the Hamiltonian is independent of time, then energy is conserved, if $H = H(t)$ explicitly, then energy is not conserved.

Compare: $H = p^2/2m - qE_0x$

with: $H' = (1/2m)(p' + qE_0t)^2$

In general, if a particular Hamiltonian is independent of some generalized coordinate x, the conjugate momentum p is conserved; if the generating function for a gauge transformation introduces a dependence on x, then p is not conserved.

**That is, introducing a gauge transformation
makes it possible to alter conservation conditions.**

NEWTONIAN MECHANICS DIFFERS FROM OTHER FORMULATIONS

It has just been seen that symmetries of the Hamiltonian (or Lagrangian) function provides information about conservation conditions.

H or L are **_system functions_** containing information that is not present in Newtonian mechanics; these system functions depend on potentials.

System functions cannot be written directly in terms of fields unless there is resort to nonlocal formalisms.

Conclusion: Potentials contain more information than do the fields.

Electrostatics can be formulated self-consistently with scalar potentials alone. Any gauge transformation will preserve the particle trajectory, but will introduce some unphysical feature such as the failure of energy conservation. Implication: there exists a <u>gauge</u> that is entirely <u>physical</u>.
For electrostatics, the physical gauge is the gauge with only a scalar potential and no vector potential.

$$\phi \neq 0, \ \ \boldsymbol{A} = \boldsymbol{0}$$

Static (and quasistatic) fields are called **longitudinal fields**.
The electric field follows field lines from one plate of a parallel plate capacitor to the other. This can be viewed as the "propagation" of the electric field in a direction parallel to the polarization direction.

Any gauge transformation from the physical gauge introduces some feature(s) that are nonphysical,
even though <u>*gauge invariance holds true.*</u>

CAUTIONS ABOUT THE "PHYSICAL GAUGE" CONCEPT

The concept is: every gauge transformation induces some alteration in physical interpretation; therefore, there can be only one gauge that matches laboratory conditions in every respect.

This concept has been criticized on the grounds of an insufficiently precise definition of the "physical gauge".

So far, the debate has engaged only a small number of qualified people. The long-term outcome is still in doubt.

Interim advice: unless you relish tough debates, avoid this concept for now. (Personally, I think it is a very useful idea, and I believe the issue was actually settled long ago: the physical gauge is the Coulomb gauge.)

PLANE WAVES

A plane wave is a pure _transverse field_.
The electric field vector is perpendicular to the propagation direction.

A true plane wave has also a magnetic field **B**, equal in magnitude to the electric field **E** (Gaussian units) with both fields forming a mutually orthogonal triad with the propagation vector **k**.

This is a **vector field** that requires a vector potential for its description.

Long experience in quantum electrodynamics, quantum field theory, elementary particle physics,... has shown that the
PHYSICAL GAUGE IS THE RADIATION GAUGE:

$$\phi = 0, \quad \mathbf{A} \neq \mathbf{0}$$

The physical gauge for a mixture of static and plane wave fields has been known for many years.

It is the **Coulomb gauge**.
For present purposes it can be defined as that gauge wherein static electric fields are represented by scalar potentials and plane wave fields are represented by vector potentials.

For an atomic electron in a laser field:

$$H = \frac{1}{2}\left(\hat{p} + \frac{1}{c}\vec{A}\right)^2 + V(r)$$

in atomic units *(ℏ=m=|e|=1),* where *V(r)* is the binding potential.
In the dipole approximation, where **A** = **A(t)**, this gauge is often called the **velocity gauge**.

As laser capabilities have mushroomed in recent years, strong-field laser physics has become a field with large numbers of participants.

The field is characterized by the near-universal realization that perturbation theory does not work; and by <u>near-universal misconceptions about gauges.</u>

From a theory standpoint, there are 3 distinct approaches:
• Direct numerical solution of the Schrödinger equation.
• Analytical approximations in the length gauge.
• Analytical approximations in the velocity gauge.
With a few exceptions (significant but largely unexploited), all theoretical work is nonrelativistic and in the dipole approximation. Atomic units are used almost universally.

$$\hbar = 1; \; m_e = 1; \; |e| = 1, \; or \; q_e = -1$$

$$\Rightarrow c = \frac{1}{\alpha} = \frac{\hbar c}{e^2} \approx 137.$$

DIPOLE APPROXIMATION

Rather than dealing formally with a multipole expansion, the dipole approximation is regarded as implying 2 properties:

$$\omega t - \boldsymbol{k} \cdot \boldsymbol{r} \rightarrow \omega t$$

$$\boldsymbol{B} = 0$$

This leads to two forms of the Schrödinger equation, where *V(r)* is the atomic binding potential:

$$VG: \quad i\partial_t \Psi = \left[\frac{1}{2} \left(-i\nabla + \mathbf{A}\left(t\right) \right)^2 + V\left(r\right) \right] \Psi,$$

$$LG: \quad i\partial_t \Psi = \left[\frac{1}{2} \left(-i\nabla \right)^2 + V\left(r\right) + \mathbf{E}\left(t\right) \cdot \mathbf{r} \right] \Psi.$$

LENGTH GAUGE

There is a gauge transformation that replaces the vector potential for the applied field by the scalar potential for quasistatic electric field. (Maria Göppert-Mayer, 1931)

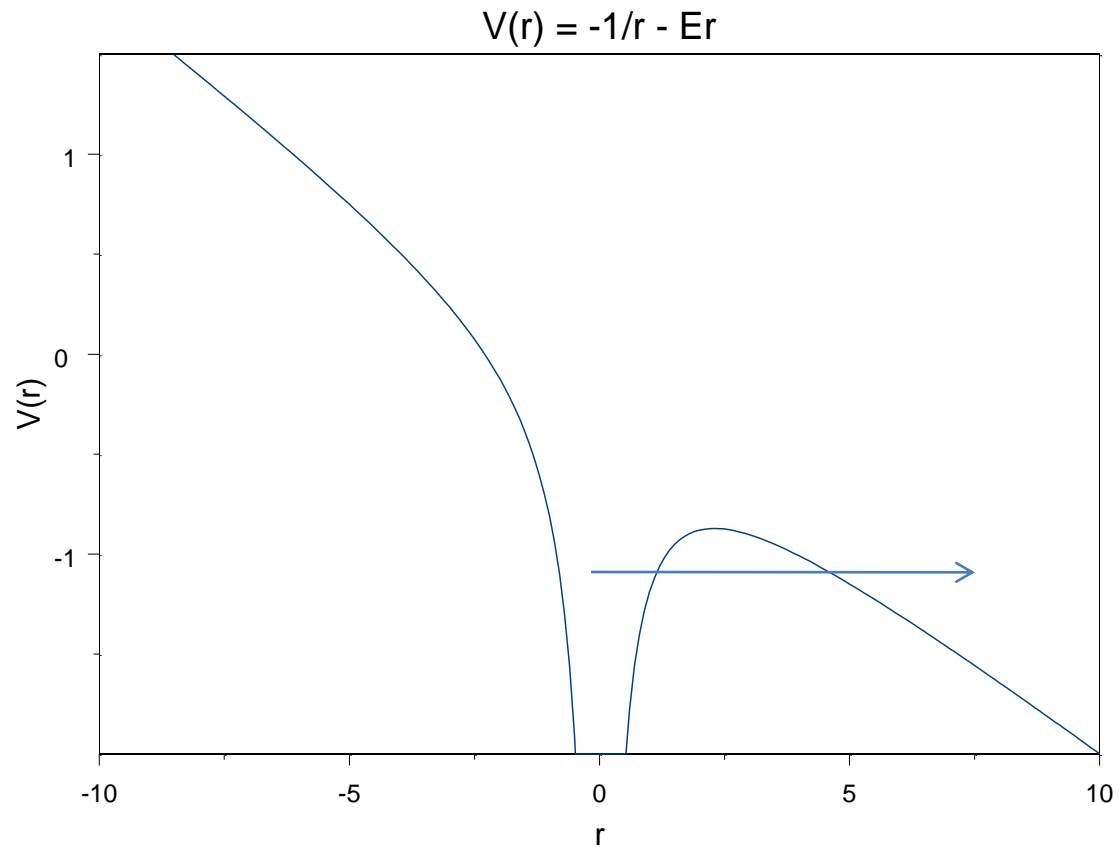$$H = \frac{1}{2}\hat{p}^2 + V(r) + \vec{r}\cdot\vec{E}(t)$$

This is a convenient gauge because both the Coulomb and the external fields are represented by scalar potentials that can simply be added.

It is the foundation for the tunneling theory of strong-field ionization. (Keldysh, Nikishov & Ritus, PPT, ADK)

Starting with a 1952 paper by Lamb, the length gauge has been regarded by many people in the AMO community as the "correct gauge" for the expression of atomic and molecular problems.

The length gauge (LG) has been favored since both the atomic binding potential and the laser field potential are scalars that can be added directly to provide a picture of the laser field "tilting" the Coulomb attractive field to result in a finite potential barrier that can be penetrated by the bound electron in a _quantum tunneling_ event.
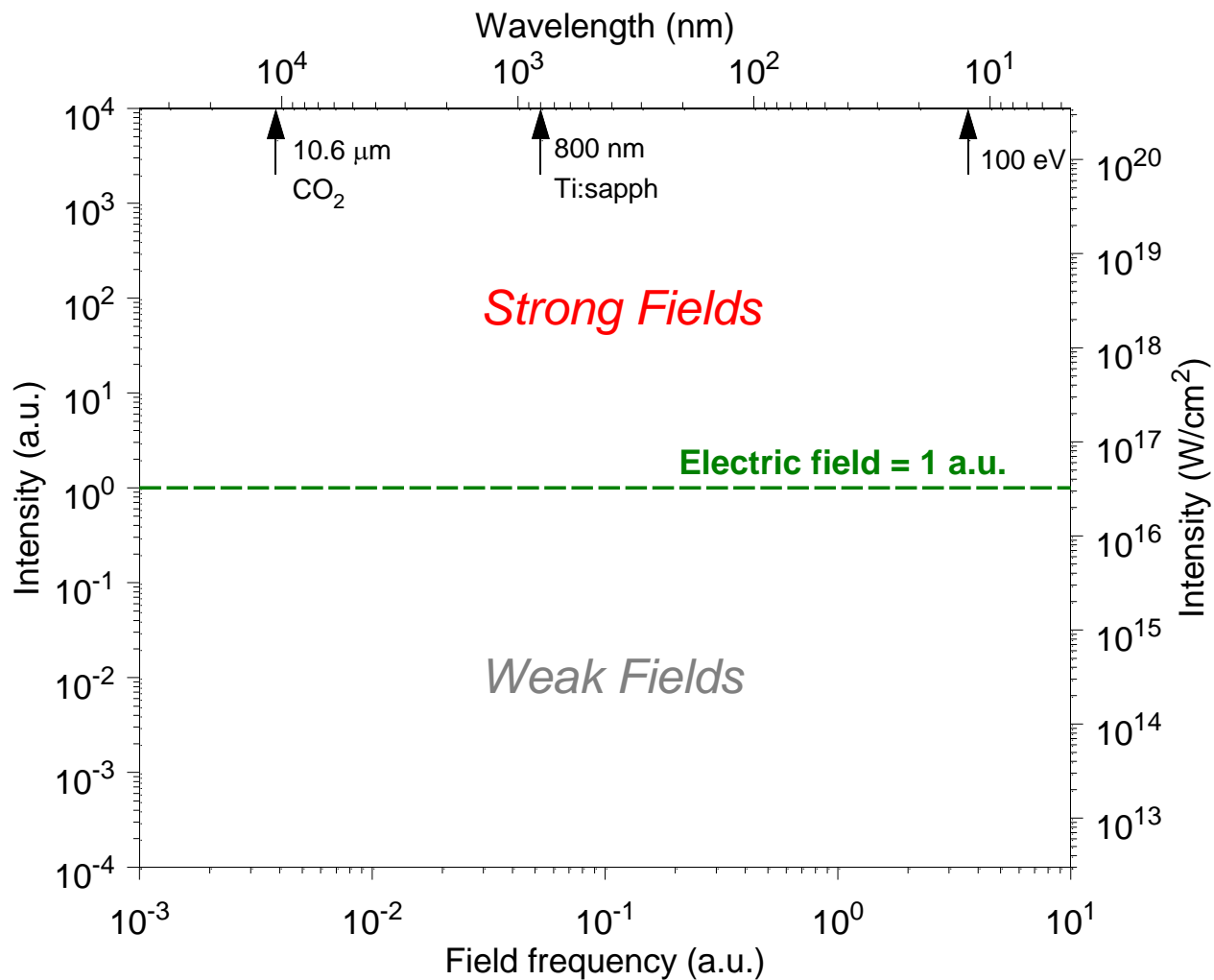

V(r) = -1/r - Er

The interaction Hamiltonian in the LG is simply $r \cdot E\,(t)$.

This leads to simple implications.

The fact that this H_I connects continuously to the static-electric-field case has been regarded not merely as an advantage, but has been treated as a **requirement.**

LASER PHYSICS VIEWED FROM THE LENGTH GAUGE: FIELD FREQUENCY

The boundary between high and low frequency should be taken to be E_B .
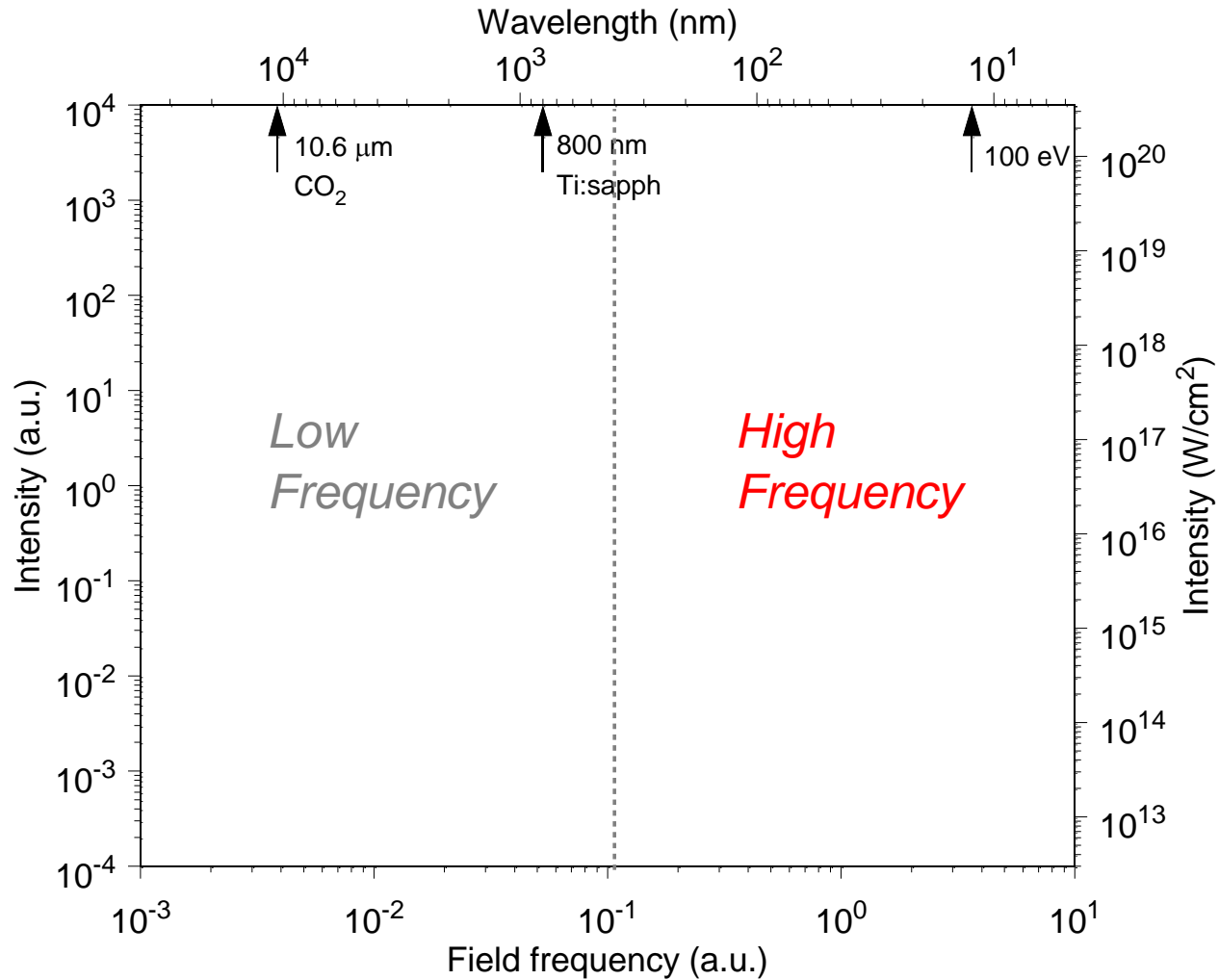
QUALITATIVE IMPLICATIONS OF THE VELOCITY GAUGE

The solution of the Schrodinger equation in the VG connects continuously to the solution for an electron in a plane-wave field from the Dirac equation.

This provides direct information on conditions for:
- The boundary for the onset of non-dipole conditions: v/c effects.
- The boundary for the onset of relativistic conditions: $(v/c)^2$ effects.

Important: The dipole approximation **<u>fails at high frequencies</u>** where $\boldsymbol{k \cdot r}$ cannot be neglected, but it **<u>also fails at low frequencies</u>** where the coupling between electric and magnetic fields becomes important. This coupling can be visualized in terms of the motion of a free electron in a plane-wave field.

Footnote: This was a surprise for the Joint Theoretical Physics Colloquium. They knew only of $\boldsymbol{k \cdot r}$ being neglected. *That* was a surprise to me.

CLASSICAL FREE ELECTRON IN A PLANE WAVE FIELD

Electron follows a figure-8 trajectory resulting from the combined action of the electric and magnetic fields.
α_0 = amplitude along electric field direction
β_0 = amplitude along direction of propagation
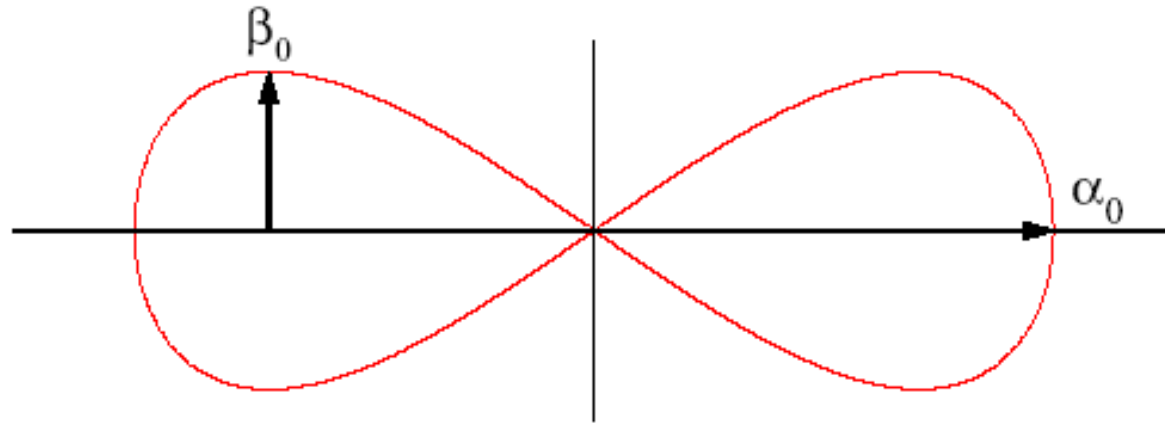
$$\alpha_0 = \frac{c}{\omega}\sqrt{2z_f}, \quad \beta_0 = \frac{c}{\omega}\frac{z_f}{4}, \quad \frac{\beta_0}{\alpha_0} = \frac{1}{4}\sqrt{\frac{z_f}{2}}, \quad z_f \equiv \frac{2U_p}{mc^2}$$

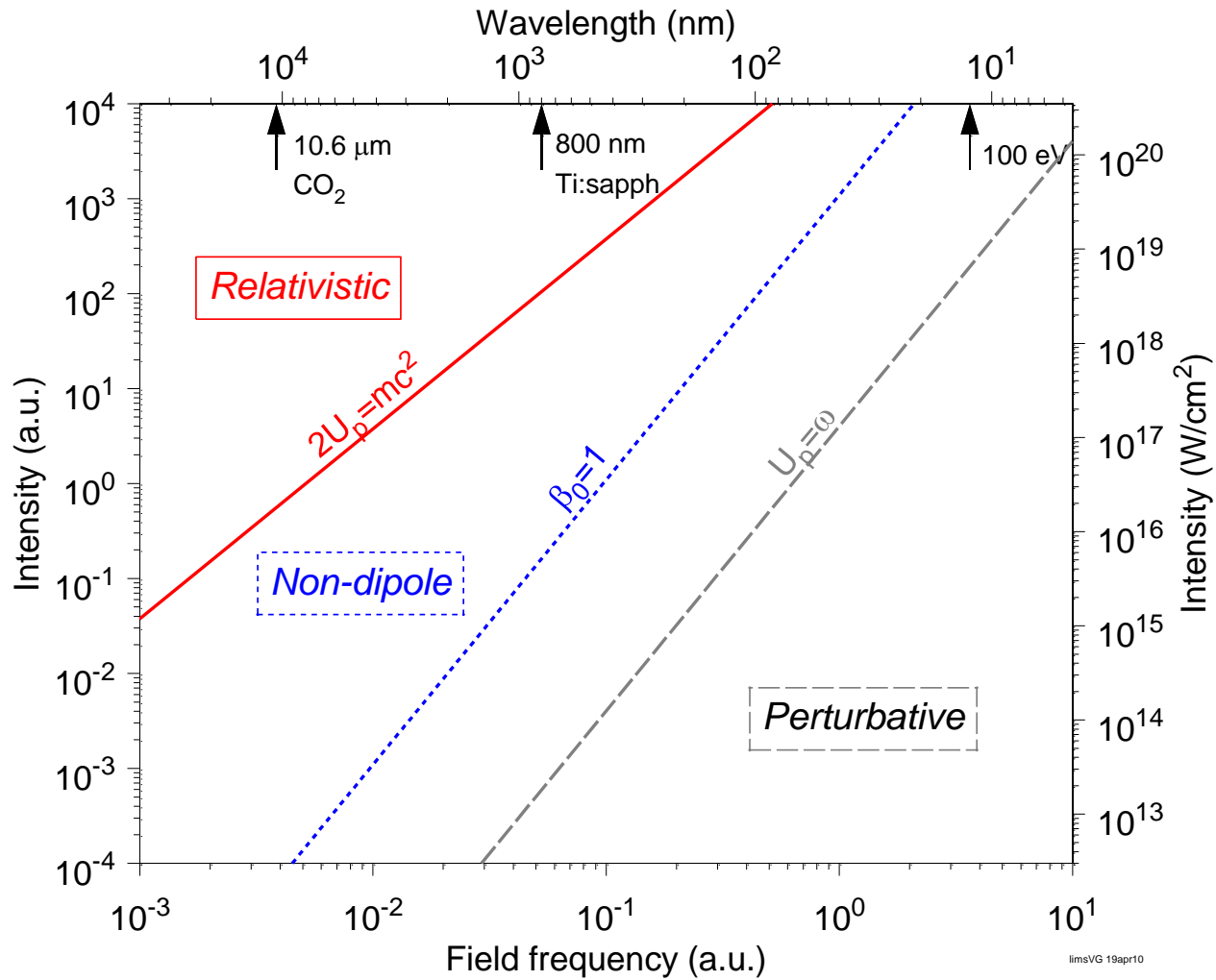## DOMAINS FOR AN ELECTRON IN A PLANE-WAVE FIELD

Lasers produce plane-wave (PW) fields, not quasistatic (QSE) fields, so the beyond-dipole-approximation results in the VG have significance that the LG does not have.

Specifically, the limits of relativistic effects and magnetic field effects shown on the next slide have real laboratory significance.

Also shown is the upper limit on intensity for which perturbation theory is applicable. This follows from both a relativistic investigation [HRR, J. Math. Phys. **3**, 387 (1962)] and a nonrelativistic investigation [HRR, Phys. Rev. A **22**, 1786 (1980)].

Most of what follows here was a surprise for the Joint Theoretical Physics Colloquium, including intensity-dependent failure of the dipole approximation.

It has just been seen that VG and LG behaviors are completely different when a broad view is taken.

Lasers produce plane waves, so the Coulomb gauge (VG in the dipole limit) is the physical gauge.
However, because of gauge invariance, the LG can nevertheless give correct results for calculation of explicit laboratory-measurable quantities.

The domain in which the LG (and all tunneling theories) can provide useful information is shown by the shaded area in the following slide from HRR, Phys. Rev. Lett **101**, 043002 (2008).

Footnote: An important property of this limited zone for the usefulness of tunneling methods is that a tunneling analysis in itself gives no hint of its own limited applicability.
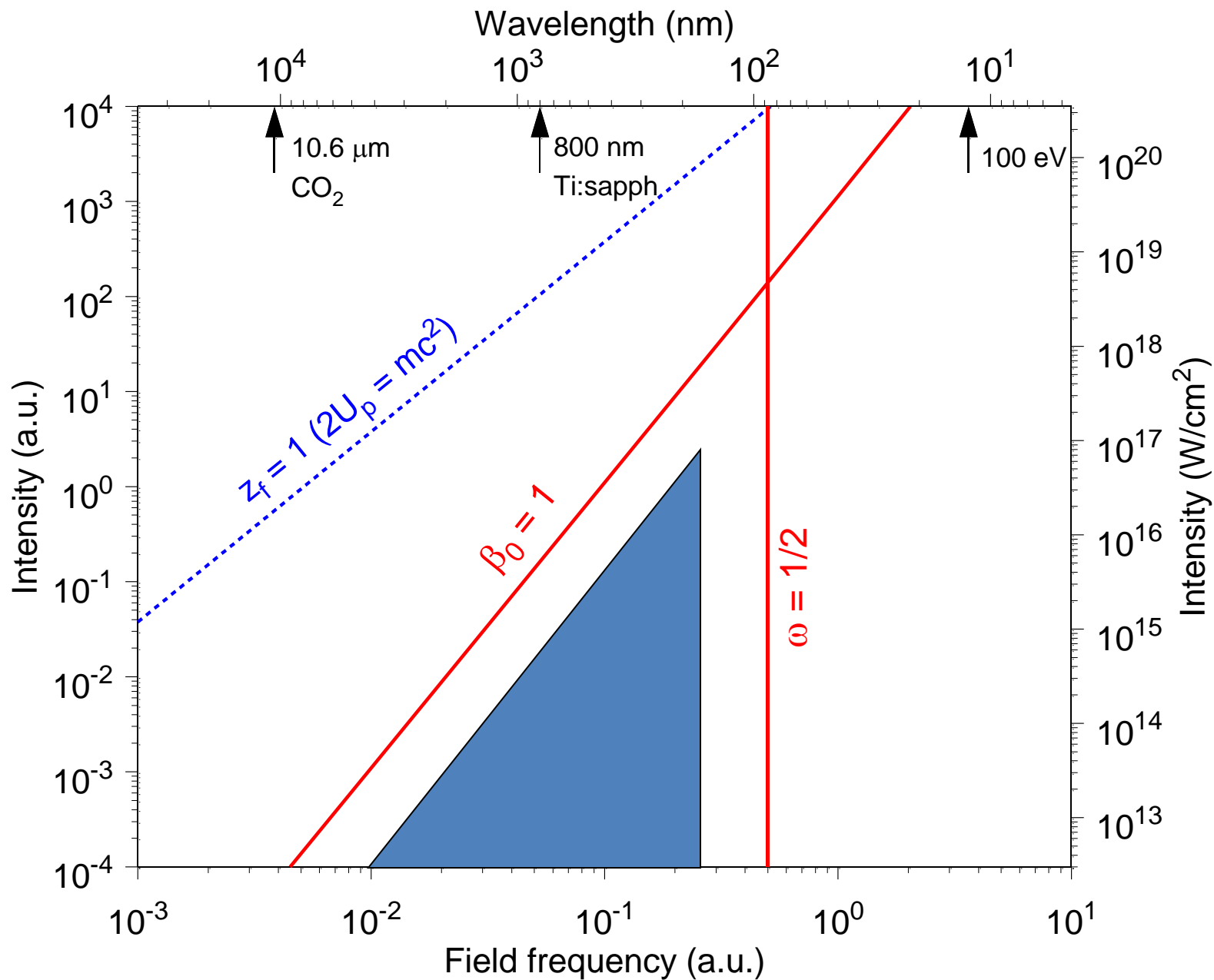
Much success has been achieved with the length gauge in general and tunneling ionization in particular, but according to what has been shown, the length gauge contradicts the result that the **physical gauge** is that gauge where transverse fields (laser fields) are represented by vector potentials **A**.

Since tunneling ionization employs a non-physical gauge, "laboratory measurables" will be predicted correctly, but there should exist physical inconsistencies such as the failure of energy conservation in the classical static field example.

What are the contradictions associated with tunneling?

INCORRECT PREDICTIONS FROM TUNNELING THEORY

• A laser field is strong when the electric field is of the order of one atomic unit; that is, when the external field is of the order of magnitude of internal fields in the atom. (Already examined.)

• A single parameter – the Keldysh parameter γ – is sufficient for scaling.

• Weak fields can be characterized as in the *multiphoton domain* and strong fields are in the *tunneling domain*.

• A *tunneling limit* or *classical limit* will be approached at long wavelengths of the field.

• There are no evident upper and lower frequency limits on the dipole approximation.

• Higher harmonics are not produced by circularly polarized fields because the photoelectron "walks away" from the atom.

A charged particle in a plane-wave field has a minimal interaction energy with the field. That is, even a particle at rest on average has a minimal interaction energy with the field (sometimes called "quiver energy"). This is the ponderomotive energy U_p .

$$U_p = \frac{I}{4\omega^2}, \qquad I = field\ intensity, \qquad \omega = field\ frequency$$

KELDYSH GAMMA PARAMETER

$$\gamma = \frac{tunneling\ time}{wave\ period}$$

$$= \sqrt{\frac{E_B}{2U_p}} = \frac{\omega}{I^{1/2}}\sqrt{2E_B}$$

Based on the tunneling concept, the following classification is nearly universal in the strong-field community:

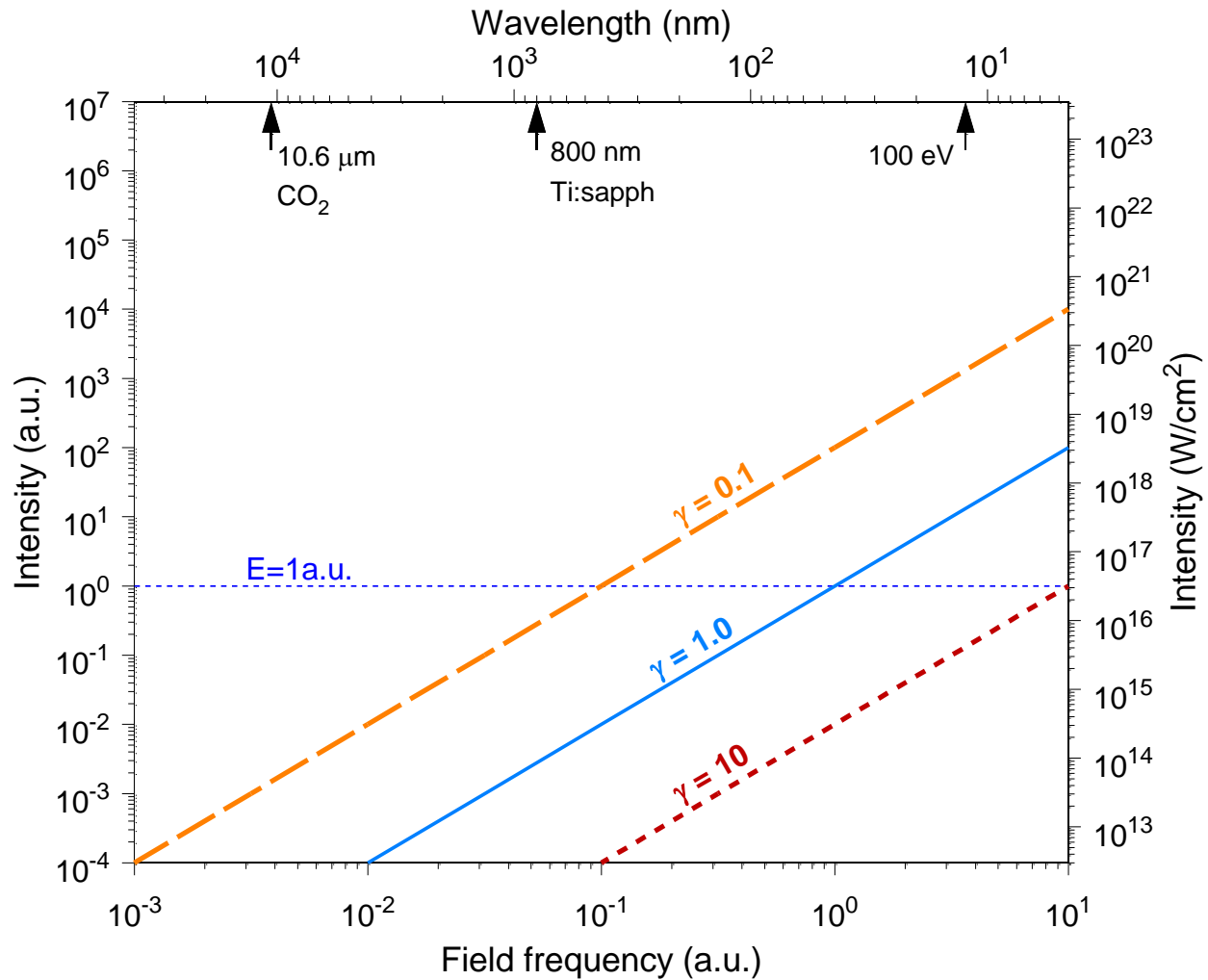$$\gamma = \begin{cases} < 1, & tunneling\ domain \\ \approx 1, & transition\ region \\ > 1, & multiphoton\ domain \end{cases}$$

35

If the binding energy is set at $E_B = 0.5$ *a.u.* (ground-state hydrogen), then

$$I = (\omega / \gamma)^2$$

Plotted as I *vs.* ω, this produces the following simple diagram that completely contradicts the "standard tunneling / multiphoton" classification.

A fixed value of γ cannot distinguish "tunneling" from "multiphoton". $\gamma = 0.1$ is supposedly "tunneling", but the right-hand end of this line is at *10 a.u. = 272 eV.*

The limit $\omega \to 0$ is regarded as the "tunneling limit", which is identified as the static limit, which is unquestionably tunneling – if that were the correct limit.

However, the next figure shows that, viewed as a plane wave (i.e. a laser field), $\omega \to 0$ is actually an extreme relativistic limit.
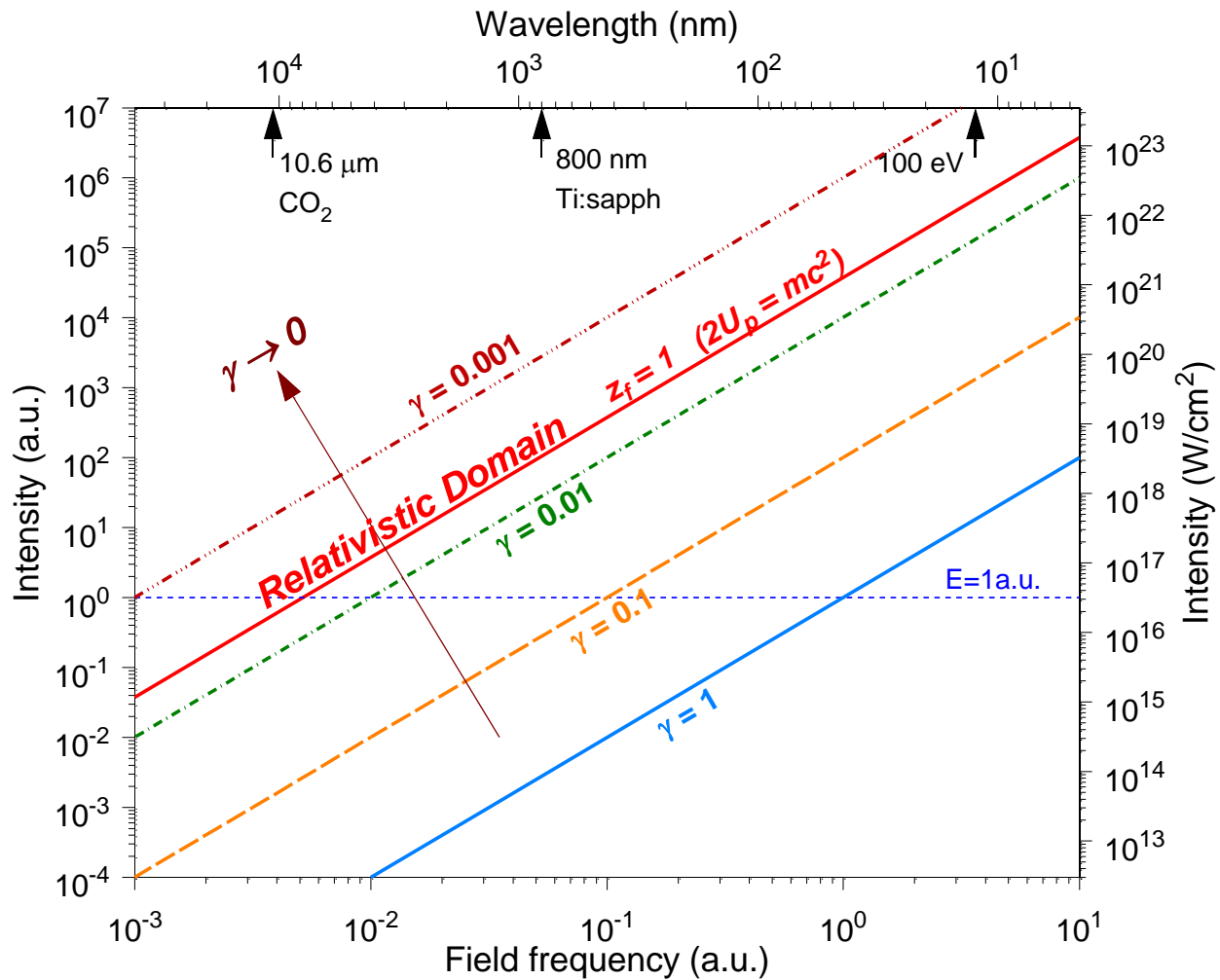
The line at $2U_p = mc^2$ is where the minimum interaction energy of an electron with the laser field is of the order of the rest energy of the electron. The left / upper corner of the graph is an extreme relativistic zone. Laser-induced ionization *__is not a static limit.__*

THE CAUSE OF THIS FUNDAMENTAL PROBLEM

The length gauge is **_NOT_** the physical gauge for laser-induced phenomena.

The Coulomb gauge **_IS_** the physical gauge.

The tunneling picture (LG picture) is completely erroneous when used outside the limited domain where there is a gauge equivalence to the Coulomb gauge (VG).

Within the gauge equivalence domain, the LG can give correct answers for physical measurables, but its qualitative inferences are seriously in error.

This simple error has led astray an entire community of very smart people.

The ponderomotive energy U_p is a fundamental quantity in strong-field physics. (It does not appear in any finite order of perturbation theory – but that is a different interesting story.)

The ponderomotive energy can be such that

$$U_p \gg E_B$$

even within the domain where the dipole approximation is valid.
In atomic units,

$$U_p = \frac{1}{2}\left\langle \frac{A}{c} \right\rangle^2$$

where the angle bracket means a time average over a full period. Within the dipole approximation,

$$A^2 = A^2(t) \quad \Rightarrow \quad U_p = U_p(t)$$

The problem is that U_p is absolutely basic in strong fields, it can be large, but there is an elementary theorem in classical mechanics that any quantity that is solely a function of time can be removed from the Hamiltonian without consequences. (This is obvious from the Hamilton equations of motion.)

If U_p is removed, strong-field physics gives wrong answers. What is going on?

The answer to this paradox can be found in a classical – quantum distinction. In classical physics, measurements are direct; in quantum physics, measurements are not made inside the interaction region, they are made outside that region.

That means that measurements are indirect in quantum mechanics; if an *f(t)* function is removed inside the region it causes an unphysical shift with respect to the measuring instruments. If that shift is a function of the vector potential, it is meaningless to remove it for the measuring instruments, which know nothing of laser fields. Hence, it cannot be removed within the interaction region.

Basic requirement: The transition-causing interaction occurs only within a domain bounded in space and time.

(Example: transitions in the focus of a pulsed laser.)

Important properties of an S matrix as derived here:
- ➢ *It can be formulated entirely in terms of quantities measurable in the laboratory.*
- ➢ *It does not require that dynamics be tracked over time. Equations of motion are incorporated in the S matrix.*
- ➢ *There is no need for "adiabatic decoupling".*
- ➢ *It gives unambiguous rules for gauge transformations.*
- ➢ *It can be applied to any process as long as the space-time domain of the interaction region is bounded.*

The general problem is evoked of an atomic electron subjected to a pulsed, focused laser beam. This is a convenience, not a requirement.

There will be a complete set of states $\{\Psi_n\}$ that satisfy the Schrödinger equation describing the atomic electron that may be undisturbed or in interaction with a laser beam:

$$i\partial_t \Psi = H\Psi$$

The outcome of any experiment will be measured by laboratory instruments that never experience a laser field. As far as the laboratory instruments are concerned, there is a complete set of states $\{\Phi_n\}$ that satisfy the Schrödinger equation describing an atomic electron that does **NOT** experience the laser field:

$$i\partial_t \Phi = H_O\Phi$$

By hypothesis, the laser pulse is finite, so

$$\lim_{t \to \pm\infty} \left[H(t) - H_0 \right] = 0$$

We can organize the two complete sets of Φ states and Ψ states so that they correspond at $t \to -\infty$:

$$\lim_{t \to -\infty} \left[\Phi_n(t) - \Psi_n(t) \right] = 0$$

After the laser interaction has occurred, the only way for the laboratory instruments to discover what has happened is to form overlaps of all possible final Φ_f states with the state that began as a particular Ψ_i state. This is the S matrix:

$$S_{fi} = \lim_{t \to +\infty} \left(\Phi_f, \Psi_i \right)$$

Subtract the amplitude that no transition has occurred:

$$M_{fi} \equiv (S-1)_{fi} = \lim_{t \to +\infty} \left(\Phi_f, \Psi_i \right) - \lim_{t \to -\infty} \left(\Phi_f, \Psi_i \right)$$

We now have the form of a perfect differential:

$$M_{fi} = \int_{-\infty}^{+\infty} dt \frac{\partial}{\partial t}(\Phi_f, \Psi_i)$$

$$M_{fi} = \int_{-\infty}^{+\infty} dt [(\partial_t \Phi_f, \Psi_i) + (\Phi_f, \partial_t \Psi_i)]$$

$$i\partial_t \Phi = H_0 \Phi \Rightarrow \partial_t \Phi = -iH_0 \Phi$$

$$i\partial_t \Psi = (H_0 + H_I)\Psi \Rightarrow \partial_t \Psi = -i(H_0 + H_I)\Psi$$

$$M_{fi} = \int_{-\infty}^{+\infty} dt [(-iH_0 \Phi_f, \Psi_i) + (\Phi_f, -i(H_0 + H_I)\Psi_i)]$$

$$M_{fi} = \int_{-\infty}^{+\infty} dt [i(\Phi_f, H_0 \Psi_i) - i(\Phi_f, (H_0 + H_I)\Psi_i)]$$

$$M_{fi} = -i\int_{-\infty}^{+\infty} dt (\Phi_f, H_I \Psi_i)$$

An alternative form is especially useful for strong-field problems. Instead of making a one-to-one correspondence of Φ and Ψ states at $t \rightarrow -\infty$, do it at $t \rightarrow +\infty$ and then look for the probabilities that particular initial states could have led to this final result.

The end result of this alternative procedure is the transition matrix element

$$M_{fi} = -i \int_{-\infty}^{\infty} dt \left(\Psi_f, H_I \Phi_i \right)$$

The first form above is called the *direct-time S matrix*, and the second form is the *time-reversed S matrix.*

In general, Φ states are known exactly or can be simulated accurately (for example, by analytical Hartree-Fock procedures). It is the Ψ states that present the problem because they represent conditions for an electron simultaneously subjected to the Coulomb attraction and to the laser field. No exact solutions are known.

If the initial state Ψ_i is to be approximated, this is very difficult because the laser field (by hypothesis) is too strong to be treated perturbatively, and so is the Coulomb field because in an initial bound state the Coulomb field has a singularity at the origin.

For the final state Ψ_f, the electron is unbound, and the laser field can be assumed to be more important than the Coulomb field. This identifies the time-reversed S matrix as the preferred form, and it will be employed exclusively hereafter.

It is widely known and easily shown that a gauge transformation applied to the Schrödinger equation gives the result

$$(H - i\partial_t)' = U(H - i\partial_t)U^{-1}$$

where U is the unitary operator that generates the gauge transformation. Although it is true that

$$\Psi' = U\Psi$$

in order to have form invariance of the Schrödinger equation, the first equation above will not give the unitary transformation of operators

$$O' = UOU^{-1}$$

if U is time-dependent.

The above is a rebuttal to the statement in the Preamble:
• A gauge transformation is a unitary transformation.

In AMO papers and textbooks, one commonly finds the statement that quantum transition amplitudes are manifestly gauge invariant. This is justified by the expression

$$(\Phi_f, H_I \Psi_i) \rightarrow \left((U\Phi_f), (U H_I U^{-1})(U\Psi_i)\right)$$
$$\rightarrow \left(\Phi'_f, H'_I \Psi'_i\right)$$

That is an expression for a change of quantum picture, not for a gauge transformation.

In a gauge transformation, field-dependent quantities change, the non-interacting reference states Φ **do not change**. In a gauge transformation

$$(\Phi_f, H_I \Psi_i) \rightarrow \left(\Phi_f, (U H_I U^{-1})(U\Psi_i)\right) = (\Phi_f, H'_I \Psi'_i)$$

Demonstrating gauge invariance is nontrivial.

The above is a rebuttal to the statement in the Preamble:
• Quantum transition amplitudes are manifestly gauge invariant.

A FUNCTION OF TIME ONLY *f(t)*, CAN BE REMOVED BY A GAUGE TRANSFORMATION

This is not the same as simply dropping a function of time. A gauge transformation should, under appropriate conditions, make it possible to calculate laboratory measurables.

The ponderomotive energy can be the cause of "channel closings", and also the cause of "high-frequency stabilization". These are physically important effects. A dilemma appears to remain.

The simplest case to consider is the stabilization matter. "Stabilization" is that property of strong-field phenomena where, beyond a certain field intensity, further increases in intensity lead to a *decline* in the transition rate rather than an increase.

The next figures [from HRR, J. Opt. Soc. Am. B **13**, 355 (1996)] show the results of calculating with the SFA (Strong-Field Approximation) and the HFA (High Frequency Approximation). The stabilization occurs at about the same intensity with both methods, but $A^2(t)$ is included in the SFA, and it has been removed by a gauge transformation in the HFA.
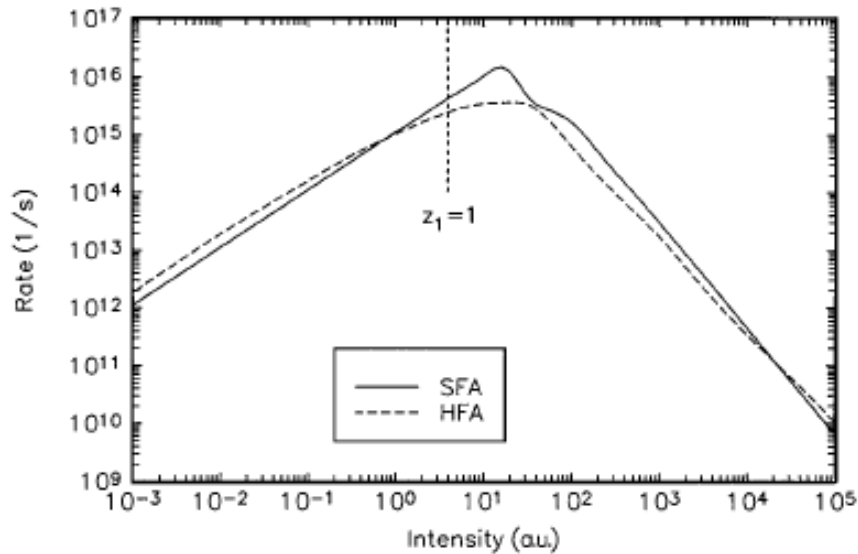
Fig. 1. Transition rates for photoionization from the ground state of hydrogen by a laser field of frequency $\omega = 2$ a.u. The solid curve is the SFA (Ref. 15), and the dashed curve is the HFA.[14] The vertical line marks the intensity at which $z_1 = 1$, where z_1 is defined in Eq. (12).
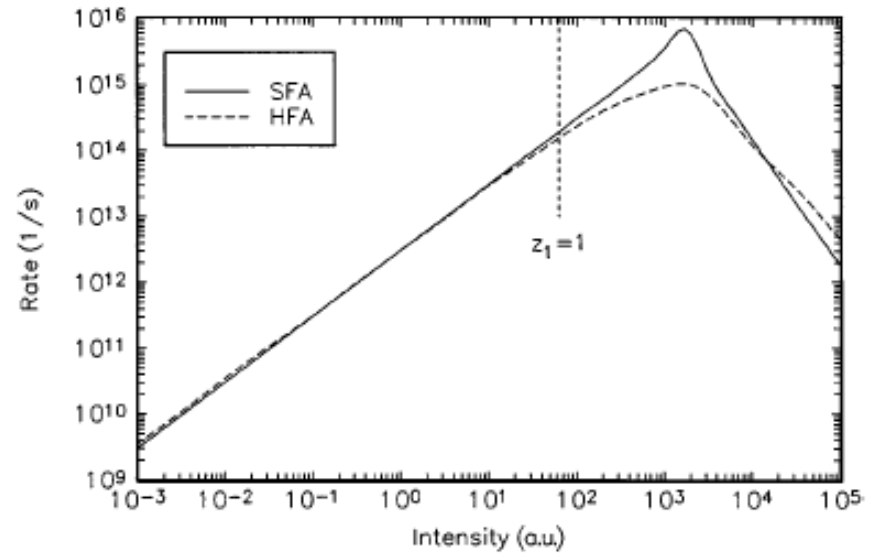
Fig. 2. Same as Fig. 1, except that $\omega = 8$.

The SFA includes $\mathbf{A}^2 (t)$; the HFA has removed that term by a gauge transformation.

Because the frequency is high ($\omega \gg E_B$), the ionization process takes place at low intensities with a single photon. In the SFA, the stabilization occurs because the ponderomotive energy has become so large as a result of increasing the field strength that the single-photon channel has closed and at least two photons are required for ionization.

There are no channel closings in the HFA because $\boldsymbol{A}^2 \ (t)$ has been removed by a gauge transformation. The physical explanation for why stabilization has occurred has changed (there is no simple explanation in the HFA), but gauge equivalence preserves the physical effect.

This is another example (beyond tunneling) where the physical explanation is dependent on the gauge even when gauge equivalence leads to the same predicted result.

• Physical processes involving electromagnetic interactions depend entirely on the electric and magnetic fields that are present.

Physical processes depend on potentials as well as fields. See the first example of the constant electric field.

• These fields can be represented as derivatives of scalar and vector potential functions that are often mathematically convenient, but are actually only auxiliary quantities.

Potentials play a more fundamental role than just being "auxiliary quantities".

• The potential functions are not unique. There are many sets of these functions corresponding to the same fields and thus to the same physical processes.

Different gauges can correspond to the same fields, but not to the same physical processes. The stabilization example is very clear on this point.

• The choice of which of the possible sets of potentials to use for a given problem is a matter of mathematical or conceptual convenience.

"Mathematical" perhaps. "Conceptual" no.

• Exactly the same considerations about the choice of gauge apply to both classical and quantum phenomena.

Quantum phenomena differ from classical in that laboratory measurements are indirect.

All of the questionable (although widely accepted) remarks in the Preamble have been addressed except the very last one:
• The most basic approach is to formulate all problems using a scalar potential. (Within the dipole approximation, this is called the *length gauge.*) If a different gauge is desired, this should be obtained through a gauge transformation from the length gauge.

This is transparently false. It is not possible to infer the 4-vector A^μ from a scalar potential ϕ. Nevertheless, there is a considerable literature bearing some well-known names that accept this assertion.