Comment on "Defining the electromagnetic potentials"

Hendrik van Hees[*]

Institut für Theoretische Physik, Goethe-Universität Frankfurt,
Max-von-Laue-Str. 1, D-60438 Frankfurt am Main, Germany
(Dated: August 29, 2020)

In this comment it is shown that the argument for a unique determination of the electromagnetic potentials in classical electrodynamics in[1] is flawed. To the contrary the "gauge freedom" of the electromagnetic potentials has proven as one of the most important properties in the development of modern physics, where local gauge invariance with its extension to non-Abelian gauge groups is a key feature in the formulation of the Standard Model of elementary particles in terms of a relativistic quantum field theory.

I. INTRODUCTION

In[1] the author claims that contrary to the standard treatment of the electromagnetic potentials in all textbooks like, e.g.,[2–4] on classical Maxwell theory the potentials are to be chosen as those of the Coulomb gauge. As shall be argued in the following, this is not only mathematically wrong but also misleading from a physical (as well as didactic) point of view since the gauge invariance of electromagnetism is the paradigmatic example for a local gauge symmetry demonstrating a general important concept for the formulation of the Standard Model of elementary particle physics, describing all hitherto observed elementary particles and their interactions in terms of a (renormalizable) relativistic quantum field theory. In this sense the claim of any fundamental a-priori preference for any specific gauge is also highly misleading from a pedagogical point of view.

From the theoretical-physics point of view it is quite commonly accepted for a long time that the fundamental laws governing the realm of classical electrodynamics are the "microscopic" Maxwell equations in differential form (for the historical context see, e.g., the remark in the introductory chapter in[2]),

$$\boldsymbol{\nabla} \times \boldsymbol{E} + \frac{1}{c}\partial_t \boldsymbol{B} = 0, \tag{1}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0, \tag{2}$$

$$\boldsymbol{\nabla} \times \boldsymbol{B} - \frac{1}{c}\partial_t \boldsymbol{E} = \frac{1}{c}\boldsymbol{j} \tag{3}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{E} = \rho, \tag{4}$$

where the Heaviside-Lorentz system of units has been used, which is more convenient for theoretical purposes than the SI units used in[1].

II. HELMHOLTZ'S THEOREM

First it is important to note that Helmholtz's theorem is applicable to time-dependent as well as to time-independent vector fields and states in a quite general form[5,6] that if a vector field $\boldsymbol{V}$ and its first derivatives, which are themselves differentiable, vanish at infinity, it can be decomposed as $\boldsymbol{V} = \boldsymbol{V}_1 + \boldsymbol{V}_2$ such that $\boldsymbol{\nabla} \times \boldsymbol{V}_1 = 0$ and $\boldsymbol{\nabla} \cdot \boldsymbol{V}_2 = 0$. With given source, $\boldsymbol{\nabla} \cdot \boldsymbol{V} = \boldsymbol{\nabla} \cdot \boldsymbol{V}_1 = J$, and curl $\boldsymbol{\nabla} \times \boldsymbol{V} = \boldsymbol{\nabla} \times \boldsymbol{V}_2 = \boldsymbol{C}$, the decomposition is unique up to additive constants for the vector fields $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$. In the following we tacitly assume the conditions on the fields needed for the following manipulations being justified.

Further there are theorems that any curl-free vector field can be written (at least in any simply connected region of space) as the gradient of a scalar potential, i.e., $\boldsymbol{V}_1 = -\boldsymbol{\nabla}\Phi$, where Φ is unique up to a constant and any source-free vector field can be written as the curl of a vector potential $\boldsymbol{V}_2 = \boldsymbol{\nabla} \times \boldsymbol{A}$, and of course $\boldsymbol{A}$ is unique only up to an arbitrary gradient field, and this freedom can be used to impose one constraint condition ("gauge condition") on $\boldsymbol{A}$.

Defining $\boldsymbol{\nabla} \cdot \boldsymbol{V} = J$ and $\boldsymbol{\nabla} \times \boldsymbol{V} = \boldsymbol{C}$ we have

$$\boldsymbol{\nabla} \cdot \boldsymbol{V} = \boldsymbol{\nabla} \cdot \boldsymbol{V}_1 = -\Delta\Phi = J, \quad \boldsymbol{\nabla} \times \boldsymbol{V} = \boldsymbol{\nabla} \times \boldsymbol{V}_2 = \boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{A}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \boldsymbol{A}) - \Delta\boldsymbol{A}. \quad (5)$$

Since we know from electrostatics how to solve the Poisson equation with the Green's function of the Laplace operator (here for "free space", i.e., without boundary conditions for Cauchy or Neumann problems as needed in electrostatics at presence of conductors or dielectrics), it is convenient to impose the additional constraint $\boldsymbol{\nabla} \cdot \boldsymbol{A} = 0$ ("Coulomb gauge condition"), such that

$$\Phi(\boldsymbol{x}) = \int_{\mathbb{R}^3} \mathrm{d}^3x' \frac{J(\boldsymbol{x}')}{4\pi|\boldsymbol{x} - \boldsymbol{x}'|}, \quad \boldsymbol{A}(\boldsymbol{x}) = \int_{\mathbb{R}^3} \mathrm{d}^3x' \frac{\boldsymbol{C}(\boldsymbol{x}')}{4\pi|\boldsymbol{x} - \boldsymbol{x}'|}. \quad (6)$$

These formulae can be proven using Green's theorem. Then one has

$$\boldsymbol{V}_1 = -\boldsymbol{\nabla}\Phi + \boldsymbol{V}_1^{(0)}, \quad \boldsymbol{V}_2 = \boldsymbol{\nabla} \times \boldsymbol{A} + \boldsymbol{V}_2^{(0)}, \quad (7)$$

where $\boldsymbol{V}_1^{(0)} = \mathrm{const}$ and $\boldsymbol{V}_2^{(0)} = \mathrm{const}$. Of course, if it is known that $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ vanish at infinity, e.g., if J and $\boldsymbol{C}$ have compact support, these constants are both determined to vanish given the potentials (6).

As we shall see, however, Helmholtz's decomposition theorem is not of prime importance to introduce the electromagnetic potentials. For this it is sufficient that a curl-free vector field can be written as the gradient of a scalar potential and that a source-free field can be written as the curl of a vector potential. For a given curl-free vector field its scalar potential is defined up to an additive constant, and for a given source-free vector field its vector potential is only determined up to a gradient of an arbitrary scalar field. In fact, as we shall see, for the solution of Maxwell's equation with given sources ρ and $\boldsymbol{j}$ the Helmholtz decomposition theorem is of not too much practical use. One rather needs a Green's function of the D'Alembert operator $\Box = \Delta - 1/c^2\partial_t^2$, of which in classical electrodynamics usually the **retarded propagator** is the relevant one (for reasons of causality).

III.   THE ELECTROMAGNETIC POTENTIALS

To see, why the claim that there is a preferred or even unique choice of the electromagnetic potentials is flawed, in this Sect. we briefly summarize the standard textbook procedure in introducing the electromagnetic potentials and arguing why they are only defined only up to a gauge transformation.

The electromagnetic potentials are introduced using the homogeneous Maxwell equations (1) and (2). Though they have profound physical meaning, from a mathematical point of view they are merely constraint conditions on the electric and magnetic fields, but nevertheless necessary to make the solutions of the complete set of the initial-value problem of Maxwell's equations unique, which describe the charge and current densities as the sources of the electromagnetic field and thus provide the dynamical equations of motion.

The homogeneous Maxwell equations (1) and (2) imply the *existence* of a vector and a scalar potential $\boldsymbol{A}$ and Φ such that

$$\boldsymbol{E} = -\vec{\nabla}\Phi - \frac{1}{c}\partial_t\boldsymbol{A}, \quad \boldsymbol{B}(t, \boldsymbol{x}) = \boldsymbol{\nabla} \times \boldsymbol{A}(t, \boldsymbol{x}). \quad (8)$$

It is also clear that the potentials are *not uniquely defined* by the electromagnetic field, $(\boldsymbol{E}, \boldsymbol{B})$ since a gauge transformation to new potentials Φ' and $\boldsymbol{A}'$,

$$\Phi' = \Phi + \frac{1}{c}\partial_t \chi, \quad \boldsymbol{A}' = \boldsymbol{A} - \boldsymbol{\nabla}\chi, \tag{9}$$

with an arbitrary scalar field χ leads to the same electromagnetic field $(\boldsymbol{E}, \boldsymbol{B})$. While $\boldsymbol{E}$ and $\boldsymbol{B}$ are observable fields, operationally defined as providing the Lorentz force $\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}/c)$, the potentials are not directly observable and only defined modulo a gauge transformation (9).

Using (8) in the inhomogeneous Maxwell equations (3) and (4) yields

$$-\Box \boldsymbol{A} + \boldsymbol{\nabla}\left(\boldsymbol{\nabla} \cdot \boldsymbol{A} + \frac{1}{c}\partial_t \Phi\right) = \frac{1}{c}\boldsymbol{j}, \tag{10}$$

$$-\Delta \Phi - \frac{1}{c}\partial_t \boldsymbol{\nabla} \cdot \boldsymbol{A} = \rho. \tag{11}$$

Here, the d'Alembert operator is used with the sign convention as in[1], i.e., $\Box = \Delta - 1/c^2\partial_t^2$. It is clear that these two equations alone do not resolve the ambiguity in the choice of the potentials since these equations are of course still gauge invariant, because they are formulated originally in terms of the Maxwell equations (3) and (4) involving only the gauge invariant fields $(\boldsymbol{E}, \boldsymbol{B})$. Thus (10) and (11) do not provide any constraint for the choice of gauge, i.e., we can still impose one constraint on the potentials to facilitate the solution of the equations (10) and (11).

A glance at (10) immediately shows that a promising choice for a gauge constraint is the **Lorenz-gauge condition**,

$$\boldsymbol{\nabla} \cdot \boldsymbol{A}_{\mathrm{L}} + \frac{1}{c}\partial_t \Phi_{\mathrm{L}} = 0. \tag{12}$$

The index L indicates the Lorenz-gauge potentials. Then from (10) and (11) one finds the inhomogeneous wave equations for the potentials

$$-\Box \Phi_{\mathrm{L}} = \rho, \quad -\Box \boldsymbol{A}_{\mathrm{L}} = \frac{1}{c}\boldsymbol{j}, \tag{13}$$

i.e., in the Lorenz gauge the equations for the Cartesian components of the vector potential decouple from each other as well as from the scalar potential.

Of course, the inhomogeneous wave equation with a given source is also not uniquely solvable but one has to impose initial as well as boundary conditions to make its solution unique, because its solutions are only determined up to a solution of the homogeneous wave equation, and this can be constrained by imposing initial conditions as well as boundary conditions. For the here discussed case of the microscopic Maxwell equations the boundary conditions are usually imposed at spacial infinity implied by the physical situation. E.g., one usually has charges and currents only in a compact spatial region and thus looks for solutions of the wave equations (12) and (13) describing waves radiating outwards from these sources. Indeed, as correctly stated in[1], also from a causality argument it is justified to choose the **retarded solution** for the potentials,

$$\Phi_{\mathrm{L}}(t, \boldsymbol{r}) = \int_{\mathbb{R}^3} \mathrm{d}^3 r' \frac{\rho(t - |\boldsymbol{r} - \boldsymbol{r}'|/c, \boldsymbol{r}')}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|}, \quad \boldsymbol{A}_{\mathrm{L}}(t, \boldsymbol{r}) = \int_{\mathbb{R}^3} \mathrm{d}^3 r' \frac{\boldsymbol{j}(t - |\boldsymbol{r} - \boldsymbol{r}'|/c, \boldsymbol{r}')}{4\pi c|\boldsymbol{r} - \boldsymbol{r}'|}. \tag{14}$$

The initial condition can then be satisfied by adding an appropriate solution of the homogeneous wave equations, $\Box \Phi_{\mathrm{L}} = 0$ and $\Box \boldsymbol{A}_{\mathrm{L}} = 0$. This of course implies that also the physical em. field $(\boldsymbol{E}, \boldsymbol{B})$ is given by retarded solutions and thus fulfill the demand of causal solutions that the observable electromagnetic field are "caused" by the presence of the charge and current densities

as sources, and the field depends at time t only on the configuration of these sources at the earlier times $t_{\text{ret}} = t - |\boldsymbol{r} - \boldsymbol{r}'|/c$.

One should note that this satisfies not only the mere causality condition but even the more strict condition of "Einstein causality" in special relativity, i.e., that an event in Minkowski spacetime can only be causally connected to events in its past light cone, which implies that no causal signals can travel faster than the speed of light in vacuum, c. As shown above the electromagnetic field propagates with this "limiting speed" of special relativity.

It is clear, though, that imposing this "causality constraint" on the potentials is not a priori necessary, since only the observable em. field $(\boldsymbol{E}, \boldsymbol{B})$ needs to be "causally connected" functionals of the sources $(\rho, \boldsymbol{j})$. Indeed, since the fields are given by derivatives of the potentials, the "causal choice" (14), using the retarded Green's function for the $\Box$-operator, as the solution of the equations (10) and (11) implies that also $(\boldsymbol{E}, \boldsymbol{B})$ are retarded solutions and thus fulfill the causality condition.

It is also important to note that (14) provide only a solution to Maxwell's equations if the Lorenz condition (11) indeed is fulfilled. A simple calculation shows that this is the case, if the continuity equation,

$$\partial_t \rho + \boldsymbol{\nabla} \cdot \boldsymbol{j} = 0, \tag{15}$$

i.e., the local form of charge conservation is fulfilled. This is anyway a necessary integrability condition for the Maxwell equations, and thus independent of the introduction of the potentials and the choice of their gauge.

One should also note that the Lorenz-gauge constraint (12) does *not* uniquely determine the potentials, since one can change the potentials by a gauge transformation with a scalar field χ according to (9), fulfilling the homogeneous wave equation,

$$\Box \chi = 0, \tag{16}$$

which has non-zero solutions (even such vanishing at spatial infinity like, e.g., the spherical wave $\chi(t, \boldsymbol{x}) = \chi \sin[k(ct - |\boldsymbol{r}|)/|\boldsymbol{r}|])$, without violating the Lorenz-gauge condition (12) for the new potentials. This arbitrariness of course is again irrelevant for the just determined causal physical fields $(\boldsymbol{E}, \boldsymbol{B})$, because these do not depend on the gauge and are given as retarded integrals over the sources $(\rho, \boldsymbol{j})$.

From this line of arguments it is already clear that the potentials do not need to be necessarily retarded solutions to fulfill the causal connection between sources and physical fields, and thus any other gauge, which may not allow for entirely retarded solutions is as justified as the Lorenz gauge.

This is of course, even in an extreme sense, illustrated by the other most commonly used gauge fixing, the **Coulomb gauge**. It is motivated by starting from (11) and observing that with imposing the constraint,

$$\boldsymbol{\nabla} \cdot \boldsymbol{A}_{\text{C}} = 0, \tag{17}$$

one decouples $\boldsymbol{A}_{\text{C}}$ from the equation for the scalar field, which now obeys a Poisson equation as in electrostatics (but of course in general with a time-dependent charge density),

$$-\Delta \Phi_{\text{C}} = \rho \tag{18}$$

with the solution

$$\Phi_{\text{C}}(t, \boldsymbol{r}) = \int_{\mathbb{R}^3} \mathrm{d}^3 r' \frac{\rho(t, \boldsymbol{r}')}{4\pi |\boldsymbol{r} - \boldsymbol{r}'|}. \tag{19}$$

This is of course still in some sense a "causal solution", as claimed in the paper, because the integrand only depends on the present time t but not on times $> t$, but it obviously seems to

violate "Einstein causality", because the scalar potential at time t is determined by the present charge configuration at this time t but gets "instantaneous contributions" from points $\boldsymbol{r}'$ which may be arbitrarily far from the observational point $\boldsymbol{r}$. As we shall see, that is not a problem at all since with the appropriate choice of solutions one finally ends up with the same retarded physical fields $(\boldsymbol{E}, \boldsymbol{B})$ as with the retarded potentials from the Lorenz-gauge potentials.

This can be seen by using (18) and the Coulomb-gauge condition (16) in (10), which gets

$$-\Box \boldsymbol{A}_{\mathrm{C}} = \frac{1}{c}\boldsymbol{j}_{\perp} \tag{20}$$

with

$$\boldsymbol{j}_{\perp}(t, \boldsymbol{x}) = \boldsymbol{j}(t, \boldsymbol{x}) - \partial_t \boldsymbol{\nabla}\Phi_{\mathrm{C}}(t, \boldsymbol{x}) = \boldsymbol{j}(t, \boldsymbol{x}) - \boldsymbol{\nabla}\int_{\mathbb{R}^3} \mathrm{d}^3 r' \frac{\partial_t \rho(t, \boldsymbol{r}')}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|}. \tag{21}$$

To see that (20) is consistent with the Coulomb-gauge condition (17), i.e., with $\boldsymbol{\nabla} \cdot \boldsymbol{j}_{\perp}$ we again need the continuity equation (15) to rewrite (21) to

$$\boldsymbol{j}_{\perp}(t, \boldsymbol{x}) = \boldsymbol{j}(t, \boldsymbol{x}) - \boldsymbol{\nabla}\int_{\mathbb{R}^3} \mathrm{d}^3 r' \frac{-\boldsymbol{\nabla}' \cdot \boldsymbol{j}(t, \boldsymbol{r}')}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|}. \tag{22}$$

Now taking the divergence of this equation indeed gives

$$\boldsymbol{\nabla} \cdot \boldsymbol{j}_{\perp}(t, \boldsymbol{x}) = \boldsymbol{\nabla} \cdot \boldsymbol{j}(t, \boldsymbol{x}) - \Delta \int_{\mathbb{R}^3} \mathrm{d}^3 r' \frac{-\boldsymbol{\nabla}' \cdot \boldsymbol{j}(t, \boldsymbol{r}')}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|} = 0. \tag{23}$$

To get retarded solutions for the fields, it seems to be appropriate to solve (20) with the retarded propagator, i.e.,

$$\boldsymbol{A}_{\mathrm{C}}(t, \boldsymbol{r}) = \int_{\mathbb{R}^3} \mathrm{d}^3 r' \frac{\boldsymbol{j}_{\perp}(t_{\mathrm{ret}}, \boldsymbol{r}')}{4\pi c|\boldsymbol{r} - \boldsymbol{r}'|}. \tag{24}$$

In the following we like to show that indeed the Coulomb-gauge potentials can be written as a gauge transformation of the retarded Lorenz-gauge potentials, which of course implies that the physical fields are the same retarded fields as derived using the Lorenz-gauge potentials.

For this proof it is convenient to introduce the vector field,

$$\boldsymbol{j}_{\parallel}(t, \boldsymbol{x}) = \boldsymbol{j}(t, \boldsymbol{x}) - \boldsymbol{j}_{\perp}(t, \boldsymbol{x}) = \boldsymbol{\nabla}\partial_t \Phi_{\mathrm{C}}(t, \boldsymbol{x}) = \boldsymbol{\nabla}\partial_t \int_{\mathbb{R}^3} \mathrm{d}^3 r' \frac{\rho(t, \boldsymbol{r}')}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|}, \tag{25}$$

where we have used (21). Then we can write (24) in the form

$$\boldsymbol{A}_{\mathrm{C}}(t, \boldsymbol{r}) = \boldsymbol{A}_{\mathrm{L}}(t, \boldsymbol{r}) - \int_{\mathbb{R}^3} \mathrm{d}^3 r' \int_{\mathbb{R}} \mathrm{d}t' \frac{\delta(t' - t + |\boldsymbol{r} - \boldsymbol{r}'|/c)}{4\pi c|\boldsymbol{r} - \boldsymbol{r}'|} \boldsymbol{\nabla}'\partial_{t'}\Phi_C(t', \boldsymbol{r}'). \tag{26}$$

Integration by parts yields

$$\begin{aligned}
\boldsymbol{A}_{\mathrm{C}}(t, \boldsymbol{r}) &= \boldsymbol{A}_{\mathrm{L}}(t, \boldsymbol{r}) - \int_{\mathbb{R}^3} \mathrm{d}^3 r' \int_{\mathbb{R}} \mathrm{d}t' \Phi_C(t', \boldsymbol{r}')\boldsymbol{\nabla}'\partial_{t'}\frac{\delta(t' - t + |\boldsymbol{r} - \boldsymbol{r}'|/c)}{4\pi c|\boldsymbol{r} - \boldsymbol{r}'|} \\
&= \boldsymbol{A}_{\mathrm{L}}(t, \boldsymbol{r}) - \frac{1}{c}\partial_t \boldsymbol{\nabla} \underbrace{\int_{\mathbb{R}^3} \mathrm{d}^3 r' \int_{\mathbb{R}} \mathrm{d}t' \Phi_C(t', \boldsymbol{r}')\frac{\delta(t' - t + |\boldsymbol{r} - \boldsymbol{r}'|/c)}{4\pi c|\boldsymbol{r} - \boldsymbol{r}'|}}_{\Psi_{\mathrm{CL}}(t, \boldsymbol{r})}.
\end{aligned} \tag{27}$$

With this first we have

$$\boldsymbol{A}_{\mathrm{C}}(t, \boldsymbol{r}) = \boldsymbol{A}_{\mathrm{L}}(t, \boldsymbol{r}) - \boldsymbol{\nabla}\chi_{\mathrm{CL}} \tag{28}$$

with the scalar field defining the gauge transformation from the Lorenz- to the Coulomb-gauge potentials,

$$\chi_{\mathrm{CL}} = \frac{1}{c}\partial_t \Psi_{\mathrm{CL}}(t, \boldsymbol{r}). \tag{29}$$

All we have to show to complete our proof of the gauge equivalence of the Coulomb-gauge and the Lorenz-gauge potentials is that with this definition we also fulfill

$$\Phi_{\mathrm{C}}(t, \boldsymbol{r}) = \Phi_{\mathrm{L}}(t, \boldsymbol{x}) + \frac{1}{c}\partial_t \chi_{\mathrm{CL}}. \tag{30}$$

Now

$$\frac{1}{c}\partial_t \chi_{\mathrm{CL}} = \frac{1}{c^2}\partial_t^2 \Psi_{\mathrm{CL}} = (\Delta - \Box)\Psi_{\mathrm{CL}}. \tag{31}$$

The first term is immediately calculated from the definition of Ψ_{Cl} in (27) since the defining integral is just the retarded solution of the inhomogeneous wave equation

$$\Box\Psi_{\mathrm{CL}} = -\Phi_{\mathrm{C}}. \tag{32}$$

Further we have, again using the definition of Ψ_{CL} in (27), integrating by parts, using (18) and (14)

$$
\begin{aligned}
\Delta\Psi_{\mathrm{CL}}(t, \boldsymbol{r}) &= \int_{\mathbb{R}^3} \mathrm{d}^3 r' \int_{\mathbb{R}} \mathrm{d}t' \Phi_C(t', \boldsymbol{r}')\Delta\frac{\delta(t' - t + |\boldsymbol{r} - \boldsymbol{r}'|/c)}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|} \\
&= \int_{\mathbb{R}^3} \mathrm{d}^3 r' \int_{\mathbb{R}} \mathrm{d}t' \Phi_C(t', \boldsymbol{r}')\Delta'\frac{\delta(t' - t + |\boldsymbol{r} - \boldsymbol{r}'|/c)}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|} \\
&= \int_{\mathbb{R}^3} \mathrm{d}^3 r' \int_{\mathbb{R}} \mathrm{d}t' \frac{\delta(t' - t + |\boldsymbol{r} - \boldsymbol{r}'|/c)}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|}\Delta'\Phi_C(t', \boldsymbol{r}') \\
&\overset{(18)}{=} -\int_{\mathbb{R}^3} \mathrm{d}^3 r' \int_{\mathbb{R}} \mathrm{d}t' \frac{\delta(t' - t + |\boldsymbol{r} - \boldsymbol{r}'|/c)}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|}\rho(t', \boldsymbol{r}') \\
&\overset{(14)}{=} -\Phi_{\mathrm{L}}(t, \boldsymbol{r}).
\end{aligned}
\tag{33}
$$

Using (32) and (33) in (31) indeed leads to (30), i.e., indeed the Coulomb-gauge potentials, with the choice of a retarded solution (24) of (20), are just a gauge transformation of the retarded Lorenz-gauge potentials and thus the resulting electromagnetic fields are the same retarded solutions as derived from the Lorenz-gauge potentials, again underlining the fact that two sets of em. potentials connected by a gauge transformation with an arbitrary gauge field χ describe the same physical situation.

The Lorenz-gauge potentials are in some respects more convenient to use since (a) they admit purely retarded solutions which are usually what is needed in the physical applications and thus these potentials admit a manifestly "causal connection" with the sources and (b) the Lorenz-gauge condition is manifestly covariant under Lorentz transformations since it reads $\partial_\mu A^\mu = 0$ in four-vector notation (where $(x^\mu) = (ct, \boldsymbol{x})$, and $\partial_\mu = \partial/\partial x^\mu$ are contra- and covariant four-vector components in Minkowski space).

Nevertheless in some respects the Coulomb gauge has also some advantages. Among them is that it fixes the gauge more stringently than the Lorenz-gauge condition. Indeed if we ask for special gauge transformations,

$$\boldsymbol{A}' = \boldsymbol{A}_{\mathrm{C}} - \boldsymbol{\nabla}\chi, \quad \Phi' = \Phi + \frac{1}{c}\partial_t \chi \tag{34}$$

such that the Coulomb-gauge condition still holds, this leads to

$$\boldsymbol{\nabla} \cdot \boldsymbol{A}' = \boldsymbol{\nabla} \cdot \boldsymbol{A} - \Delta\chi = -\Delta\chi \overset{!}{=} 0. \tag{35}$$

This implies that, under the constraint that the new gauge potentials vanish at spatial infinity as the retarded solutions for localized sources (i.e., sources with compact spacial support) $\chi = 0$, i.e., the Coulomb-gauge condition is more restrictive than the Lorenz-gauge condition. I.e., in this sense it provides a **complete gauge fixing** and thus is, e.g., most convenient to quantize the electromagnetic field in the canonical operator formalism.

It is of course clear that these retarded solutions for $(\boldsymbol{E}, \boldsymbol{B})$ can also be directly derived from the Maxwell equations (1-4) without first introducing the electromagnetic potentials, leading to the so-called **Jefimenko equations**, which are, of course, equivalent to the solutions provided by the retarded solution of the Lorenz-gauge potentials.

IV. THE FLAW IN DAVIS'S ARGUMENT

Given the above standard-textbook derivation of gauge invariance of classical electrodynamics it is clear that Davis's assertion of being able to *uniquely* define the electrodynamic potentials must be flawed. This becomes clear from the paper itself since on the one hand in Sect. 4 he "proves" that the Coulomb-gauge potentials are "uniquely" determined by an apparently more "rigorous" approach to the Helmholtz decomposition theorem for vector fields, while in Sect. 5 he derives the retarded Lorenz-gauge potentials.

First of all the method to introduce certain inverse operators for the Laplace operator, Δ, and the d'Alembert operator $\Box$ is mathematically correct. It boils down to define these inverse operators as integral operators with the usual free-space Green's functions, i.e.,

$$\begin{aligned}
\Delta^{-1}\phi(t, \boldsymbol{x}) &= -\int_{\mathbb{R}^3} \mathrm{d}^3x' \frac{\phi(t, \boldsymbol{x}')}{4\pi|\boldsymbol{x} - \boldsymbol{x}'|}, \\
\Box^{-1}\phi(t, \boldsymbol{x}) &= -\int_{\mathbb{R}^3} \mathrm{d}^3x' \frac{\phi(t - |\boldsymbol{x} - \boldsymbol{x}'|/c, \boldsymbol{x}')}{4\pi|\boldsymbol{x} - \boldsymbol{x}'|}.
\end{aligned} \tag{36}$$

It is also correct that the so-defined operator Δ^{-1} is unique, given the physical boundary conditions of sufficiently quickly falling functions, as is summarized in Sect. II of this comment. As detailed in Sect. III the operator $\Box^{-1}$ is not unique but the retarded propagator is chosen on grounds of the given causality arguments, i.e., by an additional temporal boundary condition.

In this sense the treatment of the potentials by Davis is mathematically correct, but it is incomplete, which leads to the wrong assertion that by using these operator methods the electromagnetic potentials are uniquely determined. In contradistinction to the physically observable electromagnetic field, $(\boldsymbol{E}, \boldsymbol{B})$, the potentials are not uniquely defined by the Maxwell equations and appropriate intial/boundary conditions, but the "ambiguity", formalized by the gauge invariance of the physical observables, is not relevant for classical electromagnetism as a complete theory of electromagnetic phenomena.

In the operator-formalism language this is seen as follows. Using (2) it follows that

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{B}) = -\Delta\boldsymbol{B} \tag{37}$$

and thus, indeed uniquely,

$$\boldsymbol{B} = -\Delta^{-1}\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{B}) = -\boldsymbol{\nabla} \times \Delta^{-1}(\boldsymbol{\nabla} \times \boldsymbol{B}). \tag{38}$$

The claim that from this the choice

$$\boldsymbol{A}_{\mathrm{C}} = -\Delta^{-1}(\boldsymbol{\nabla} \times \boldsymbol{B}) \qquad \text{(incomplete!)} \tag{39}$$

were unique is, however, flawed since, according to Helmholtz's decomposition theorem applied to the vector potential, $\boldsymbol{A}$, without specifying $\boldsymbol{\nabla} \cdot \boldsymbol{A}$ the general solution for the vector potential is

$$\boldsymbol{A} = \boldsymbol{A}_{\mathrm{C}} - \boldsymbol{\nabla}\chi, \quad \boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A} = \boldsymbol{\nabla} \times \boldsymbol{A}_{\mathrm{C}} \tag{40}$$

with an arbitrary and thus *undetermined* scalar field $\chi(t, \boldsymbol{x})$. This is precisely the unavoidable "ambiguity" in defining the potentials for a given physical situation due to gauge invariance. Of course (39) is the uniquely defined Coulomb-gauge vector potential, because it fulfills the Coulomb-gauge condition,

$$\boldsymbol{\nabla} \cdot \boldsymbol{A}_{\mathrm{C}} = -\Delta^{-1}\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \boldsymbol{B}) = 0. \tag{41}$$

Now it is also clear how to make Davis's arguments complete in the further steps. First using Faraday's Law (1) leads to

$$\vec{\nabla} \times \boldsymbol{E} + \frac{1}{c}\partial_t \boldsymbol{B} = \boldsymbol{\nabla} \times \left(\boldsymbol{E} + \frac{1}{c}\partial_t \boldsymbol{A}\right) = 0. \tag{42}$$

According to Helmholtz's theorem, this leads to the existence of a scalar potential, Φ,

$$\boldsymbol{E} + \frac{1}{c}\partial_t \boldsymbol{A} = -\boldsymbol{\nabla}\Phi. \tag{43}$$

Using (4) and (40) and taking (41) this results in

$$\boldsymbol{\nabla} \cdot \left(\boldsymbol{E} + \frac{1}{c}\partial_t \boldsymbol{A}\right) = \rho - \frac{1}{c}\partial_t \Delta\chi = -\Delta\Phi \tag{44}$$

with the unique solution

$$\Phi = -\Delta^{-1}\rho + \frac{1}{c}\partial_t\chi = \Phi_{\mathrm{C}} + \frac{1}{c}\partial_t\chi. \tag{45}$$

The conclusion is that, contrary to Davis's claims, the electromagnetic potentials are determined to be the Coulomb-gauge potentials *only* up to a gauge transformation (9).

The same line of arguments also follows when using the operator formalism of Sect. 5 of Davis's paper, employing $\Box^{-1}$, defined via the retarded propagator as stated above, instead of Δ^{-1}, leading to the general solution for the potentials given by the retarded Lorenz-gauge propagator but again *only* modulo a gauge transformation,

$$\boldsymbol{A} = \boldsymbol{A}_{\mathrm{L}} - \boldsymbol{\nabla}\chi', \quad \Phi = \Phi_{\mathrm{L}} + \frac{1}{c}\partial_t\chi' \tag{46}$$

with another undetermined scalar field, χ'. Of course the retarded Lorenz-gauge potentials in the operator prescription are given by

$$\boldsymbol{A}_{\mathrm{L}} = -\Box^{-1}\frac{1}{c}\boldsymbol{j}, \quad \Phi_{\mathrm{L}} = -\Box^{-1}\rho, \tag{47}$$

as correctly stated in Davis's paper, but again also using this strategy for solving the Maxwell equations for given sources ρ and $\boldsymbol{j}$ only leads to the determination of the potentials *only* up to a gauge transformation

V. CONCLUSION

In this comment, we have clarified that the electromagnetic potentials are *not* uniquely determined by the (relativistic) causality constraint leading to a unique choice of the potentials, neither as the retarded solutions of the wave equations for the potentials in Lorenz gauge nor as the solution of the Coulomb-gauge potentials, as falsely claimed in[1]. We have illustrated that the "ambiguity" in the choice of the potentials are mathematical facts summarizing the standard-textbook approach as well as the operator approach used in this paper. It should also be emphasized that this "ambiguity" is nothing else than gauge invariance of classical Maxwell theory and is thus irrelevant for the observable electromagnetic fields since the different electromagnetic potentials related to each other by a gauge transformation represent the same physics. The causality constraint, including the more stringent Einstein causality imposed by the relativistic spacetime structure, has to be imposed only on the physically observable fields and are not a necessary condition for the unobservable electromagnetic potentials.

This has been demonstrated by the two standard examples given in most standard textbooks, the Lorenz gauge, which leads to decoupled inhomogeneous wave equations for the scalar and the components of the vector potential, and for them thus the (Einstein) causality condition can be fulfilled by using the retarded propagator of the d'Alembert operator, leading to retarded potentials and thus also to a retarded electromagnetic field. The retardation is given by the speed of light as to be expected from a massless field like the electromagnetic field. Thus these retarded solutions obey both the causality and the more stringent Einstein causality constraints as it must be for a classical relativistic field theory.

The other "extreme choice" with regard to retardation is the Coulomb gauge, which leads to an instantaneous solution for the scalar potential and in turn to a wave equation for the vector potential with a nonlocal distribution to the source. As it turns out, both the scalar and the vector potential contain non-retarded, instantaneous contributions. However, the vector potential in the Coulomb gauge consists of the sum of the retarded Lorenz-gauge vector potential and a gradient field. This immediately implies that the magnetic field $\boldsymbol{B}$ calculated from the Coulomb-gauge potential is the same retarded field as calculated from the retarded Lorenz-gauge vector potential. Then we have demonstrated that the resulting gauge field also connects the instantaneous Coulomb-gauge scalar potential with the retarded Lorenz-gauge scalar potential in the way described by the corresponding gauge transformation such that also the electric field turns out to be the same retarded field as calculated from the Lorenz-gauge potentials.

Of course one can also solve the Maxwell equations without introducing the potentials, leading to inhomogeneous wave equations, for which the unique physical choice is the retarded solution due to the usual causality arguments given above. Here the operator formalism is indeed more convenient than the standard textbook approach in leading to *unique* solutions for the physical fields, $(\boldsymbol{E}, \boldsymbol{B})$, the so-called **Jefimenko equations**. This is an elegant demonstration of the uniqueness of solutions of the initial-value problem (together with the appropriate spatial boundary conditions) of the Maxwell equations in terms of the physical fields, which is usually shown using the energy expression for the electromagnetic field, as, e.g., in[2].

It is also interesting that one can choose a different class of gauge constraints, the so-called "velocity gauges" such that the em. potentials contain retarded contributions which, however, propagate not with the speed of light but with arbitrary speeds, even larger than the speed of light. Of course, these fields are not observable either but lead again to the same physical retarded solutions for the em. field as it must be[7].

Further it should be kept in mind that the gauge invariance of classical electrodynamics is the key to use it in a consistent way in both the semiclassical description of electromagnetic interactions in non-relativistic quantum mechanics (i.e., with the charged particles described in terms of quantum

mechanics but the electromagnetic fields kept at the classical level) or in the full relativistic local quantum field theory, i.e., QED.

Already the analysis of the unitary representations of the proper orthochronous Poincaré group for free massless spin-1 fields leads to the *necessity* of formulating it as a quantized Abelian gauge theory to avoid the appearance of continuous intrinsic polarization-like degrees of freedom, which are indeed not observed in Nature[8,9].

Contrary to some claims in the literature (see, e.g.,[10,11]) neither classical electrodynamics nor quantum theory enables a unique definition of the electromagnetic potentials, but to the contrary the gauge symmetry is a *necessary* feature for the consistency of the description of the electromagnetic interaction in the quantum realm.

Acknowledgments

[*] Electronic address: hees@itp.uni-frankfurt.de

[1] A. Davis, Eur. J. Phys. **41**, 045202 (2020), `https://doi.org/10.1088/1361-6404/ab78a6`.

[2] A. Sommerfeld, *Vorlesungen über Theoretische Physik III, Elektrodynamik* (Verlag Harri Deutsch, Frankfurt/M., 2001).

[3] J. D. Jackson, *Classical electrodynamics* (John Wiley&Sons, Hoboken, NJ, 1998), 3rd ed.

[4] D. J. Griffiths, *Introduction to electrodynamics* (Cambridge University Press, Cambridge, New York, Melbourne, New Delhi, Singapore, 2017), 4 ed., `https://doi.org/10.1017/9781108333511`.

[5] A. Sommerfeld, *Vorlesungen über Theoretische Physik II, Mechanik der deformierbaren Medien* (Verlag Harri Deutsch, Frankfurt/M., 1992).

[6] O. Blumenthal, Math. Ann. **61**, 235 (1905), `https://doi.org/10.1007/BF01457564`.

[7] J. D. Jackson, Am. J. Phys. **70**, 917 (2002), `https://doi.org/10.1119/1.1491265`.

[8] R. U. Sexl and H. K. Urbantke, *Relativity, Groups, Particles* (Springer, Wien, 2001).

[9] S. Weinberg, *The Quantum Theory of Fields*, vol. 1 (Cambridge University Press, 1995).

[10] H. R. Reiss, J. Phys. B: At. Mol. Opt. Phys. **50**, 075003 (2017), `https://doi.org/10.1088/1361-6455/aa6375`.

[11] H. R. Reiss, Phys. Rev. A **100**, 052105 (2019), `https://doi.org/10.1103/PhysRevA.100.052105`.