Journal of the Optical Society of America A
OPTICS, IMAGE SCIENCE, AND VISION

External versus induced and free versus bound electric currents and related fundamental questions of the classical electrodynamics of continuous media: discussion

VADIM A. MARKEL

Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (vmarkel@pennmedicine.upenn.edu)

Received 25 April 2018; revised 28 June 2018; accepted 10 August 2018; posted 10 August 2018 (Doc. ID 328856); published 7 September 2018

Standard textbooks on classical electrodynamics frequently operate with the notions of free and bound currents (charges). Alternative terminology of external and induced currents also exists. However, a clear physical definition of these physical objects is rarely given. The term "free current" can refer in some cases to the conductivity current, which is subject to constitutive relations in a material sample. In other cases, free current refers to the current that is completely extrinsic to a given material sample and is assumed to be known *a priori* or manipulated by the experimentalist at will. Although one can argue that all currents flowing in material media are subject to some constitutive relations, there is a clear distinction in the construction of the classical electrodynamics between the external and induced currents. The aim of this paper is to clarify this distinction while pointing out that the traditional distinction between free and bound currents is arbitrary and can be abandoned. In addition, the paper considers some relevant fundamental questions of classical electrodynamics, including the derivation of macroscopic Maxwell's equations, the properties of the external currents, and the physical interpretation of some auxiliary fields such as the field of polarization P.

https://doi.org/10.1364/JOSAA.35.001663

1. INTRODUCTION

The resurgence of interest in the foundations of classical electrodynamics can be evidenced by a number of recent papers [1–6]. In the author's opinion, significant confusion has developed in this field, especially concerning the concepts of free and bound charges and currents, and external and induced charges and currents. This terminology and the associated mathematical methods are often used inconsistently or even incorrectly. Consequently, graduate education in classical electrodynamics will benefit from a unified description and a clear understanding of the physical quantities mentioned above.

This paper is an attempt to present such a unified description. It is shown that the concepts of free charges and currents, as used frequently in the literature, are outdated and unnecessary and can be easily avoided. On the other hand, the concepts of external and induced charges and currents are indeed fundamental and should never be confused with each other or with free charges (currents). Unfortunately, free and external charges (currents) are often conflated and described in similar terms. However, these objects are generally not equivalent. The same is true for bound and induced charges (currents). Therefore, it appears to be important to draw a clear distinction.

The physical arguments presented in this paper are closely related to some fundamental questions of classical electrodynamics, such as the derivation of the macroscopic Maxwell's equations from their microscopic counterparts. Another important question is the interpretation of the auxiliary fields introduced in the macroscopic theory, such as the field of polarization **P**. The author believes that these questions deserve a discussion as well. Therefore, the paper is organized as follows.

In Section 2, the definitions of free charges (currents) appearing in various standard textbooks are discussed and the existing inconsistencies are pointed out. In Section 3, a self-consistent microscopic theory that serves as the first principles for the derivation of macroscopic Maxwell's equations is stated. In Section 4, the notion of external currents in the framework of the microscopic theory is introduced. The main point of this section is that the definition of external currents is situative rather than physical. In Section 5, the macroscopic Maxwell's equations are derived by using certain phenomenological postulates. The averaging of the microscopic fields plays no role in this derivation and, in fact, it is shown that such averaging cannot be adequately defined mathematically. The important

message of this section is a clear definition of the external and induced currents. Section 6 is central to this paper. Here the properties of external currents and the physical and mathematical consequences of these properties are discussed. The two important ideas introduced in Section 6 are (i) the requirement that external currents not overlap with the medium in which macroscopic Maxwell's equations are being solved and (ii) the observation that the model of infinite unbounded media is incomplete and, in many practical applications of the theory, it is important to include the medium boundaries into consideration. In Section 7, one common point of contention is discussed, namely, that the steady conductivity currents can be described in the same mathematical framework as all other induced currents in the medium and, in particular, that the conductivity current $\sigma\mathbf{E}$ can always be absorbed in the term $\partial\mathbf{P}/\partial t$, even in the static limit. The author does not suggest that the expression $\sigma\mathbf{E}$ is incorrect or should not be used when appropriate; rather, the message is that absorbing the conductivity currents into the term $\partial\mathbf{P}/\partial t$ does not lead to loss of generality and, at all finite frequencies, the latter approach is preferred, as it results in shorter formulas and fewer notations. Finally, Section 8 contains a brief summary.

The paper is largely methodological. The author does not suggest that the traditional way of doing and teaching electrodynamics is erroneous. Rather, the implicit ambiguity and lack of clarity of the definitions of free and induced charges and currents, combined with some conceptually difficult questions arising in the theory of electromagnetic nonlocality (discussed in detail below), have resulted recently in the construction of mathematically valid but physically inapplicable models in the theory of electromagnetic homogenization [7–17]. The physically incorrect elements of these models are (i) the assumption that an external (sometimes called "impressed") current can spatially overlap with a sample of electromagnetic medium, (ii) the assumption that this medium is infinite, and (iii) stemming from point (ii), a general disregard of the role and importance of the medium boundaries. This approach can be a source of serious mistakes in the theory of electromagnetic homogenization. Of course, whenever a physical model is constructed, one must use some physical intuition to make sure that the model is adequate to the reality. This inevitably results in a degree of arbitrariness. Indeed, physics, unlike mathematics, cannot be built axiomatically. However, the author hopes that the theoretical framework proposed below, while consistent with all accumulated knowledge, will significantly reduce the possibility of mistakes and, ultimately, will be less confusing for the students studying the subject.

The Gaussian system of units is used throughout the paper.

2. FREE CHARGES AND CURRENTS IN THE LITERATURE

Macroscopic Maxwell's equations with some source terms $\mathbf{J}_{\text{free}}$ and ρ_{free} on the right-hand side (sometimes written with the subscripts, sometimes without, but still designated as "free" charges and currents in the text) frequently appear in the literature. Usually, there is no further explanation of what exactly these free currents and charges are. The authors who write Maxwell's equations in this form probably assume that the terminology is so customary and clear that it does not require any explanation in a research paper. More often than not, the lack of definition of $\mathbf{J}_{\text{free}}$ and ρ_{free} is inconsequential since these quantities are simply not used in any way, or are later assumed to be zero. In other cases, as in the case of current-driven homogenization [7–17], the choice of free terms has far-reaching consequences. The problems with this approach have been discussed by us in detail elsewhere [18,19] and similar arguments (related to the importance of considering finite samples) have been also made by Vinogradov and Merzlikin [2] and by Merzlikin and Puzko [20]. Therefore, it would be reasonable to ask what the exact definition of these terms is. Unfortunately, standard textbooks are neither clear nor consistent on the subject. Indeed, consider the following examples:

In *Fundamentals of the Theory of Electricity* [21], Tamm gives the following definition:

> By free charges we shall mean, first, all the electric charges that can move over macroscopic distances under the influence of an electric field (electrons in metal and in a vacuum, ions in gases and electrolytes, etc.), and, second, charges brought in from outside onto the surface of dielectrics and violating their neutrality (for example, the charges of the intraionic lattice of solid dielectrics formed owing to the lack of ions of definite sign in this section of the dielectric so that the section as a whole is no longer neutral).

This definition is hardly satisfying. The first sub-category of free charge, according to Tamm, are the electrons that can participate in a steady current, i.e., the conduction-band electrons in metals. Of course this association is purely microscopic. In the macroscopic theory, the charge density is continuous and it is generally impossible to trace a single charge and tell how far it went "under the influence of an electric field." Even charge velocity is undefined in the macroscopic theory. The second sub-category introduced by Tamm is even more dubious. It is not really possible to put a label on a charge declaring that it was (unlike all other similar charges) "brought in from outside," if that history even matters. Besides, how is it possible to tell which particular charge violates the electric neutrality of a given dielectric body or some part thereof? Finally, what if the body *is* electrically neutral as a whole and also non-conducting? Does this mean that there are no free charges?

In *Electricity and Magnetism* [22], Purcell starts with a similar idea of free charges and currents in dielectrics being in some sense foreign or extraneous to a given sample, but makes a crucial remark:

> It is often useful to distinguish between the foreign charge Q and the charges that make up the dielectric itself. Over the former we have some degree of control—charge can be added to or removed from an object, such as the plate of a capacitor.

Here the important point is control. But what does it mean exactly to exercise some degree of control over charge or current? For example, consider a homogeneous macroscopic sphere irradiated by an incident plane electromagnetic wave. The problem in this case can be solved analytically and, therefore, we can calculate exactly the current and charge distributions inside the sphere. So, since we can predict these quantities

theoretically, do we exercise control over them in the just-described experiment?

Purcell, similarly to Tamm, also labels the conductivity current as free. However, in Section 10.14 of *Electricity and Magnetism*, he makes another important point, namely, that it is generally impossible to decompose unambiguously the total current density $\mathbf{J}$ into two contributions $\mathbf{J}_{\text{bound}}$ and $\mathbf{J}_{\text{free}}$ (the current densities of the bound and free charges). To examine the possibility of such a decomposition, Purcell invokes the notion of "molecular dipoles." Indeed, if the vector of polarization, $\mathbf{P}$, is the volume density of these "molecular dipole moments," and if this density can be determined unambiguously, then we can compute the bound current as $\mathbf{J}_{\text{bound}} = \partial\mathbf{P}/\partial t$, whereas the remainder of the total current can be designated as free and written in the form $\mathbf{J}_{\text{free}} = \sigma\mathbf{E}$. In this case, the two contributions to the total current can be disentangled on physical grounds. However, Purcell also notes that the "molecular dipole moments" are themselves not well-defined quantities, and so the whole distinction is dubious. The conclusion drawn by Purcell is the following:

> This example teaches us that in the real atomic world the distinction between bound charge and free charge is more or less arbitrary, and so, therefore, is the concept of polarization density $\mathbf{P}$.

If the distinction is indeed arbitrary, why is it introduced at all?

In this regard, it can be mentioned that the widespread belief that the vector of polarization, $\mathbf{P}$, is, by definition, the volume density of the electric dipole moment induced in a sample of a continuous medium is not generally correct, and that this interpretation of $\mathbf{P}$ is not really needed for anything—nothing in the macroscopic theory depends on it. This was apparently understood by Purcell, but the fact came to the fore with the development of the modern density functional theory (DFT) computational methods (see relevant reviews by Resta and Vanderbilt [23] and Spaldin [24]); the question was also considered recently without invocation of any quantum concepts by the author [25].

Having acknowledged the ambiguity, Purcell has claimed that, at least in the static limit, the distinction between free and bound currents makes sense:

> There is one rather obvious practical distinction—you can't have a *steady* bound charge current, one that goes on forever unchanged.

However, one can rather trivially write a constitutive relation for a conducting medium such that the steady current is given by the term $\partial\mathbf{P}/\partial t$. This point will be discussed in more detail in Section 7.

We can summarize that the various definitions of free currents and charges given in the literature are inconsistent and ambiguous. At the very least, all such definitions appeal to the microscopic description of matter. This is contradictory because, in the macroscopic theory, all microscopic notions such as point charges, charge velocity, etc., disappear from consideration. One can, for example, pose the following question: I have a piece of material of known shape with a measured dielectric permittivity of $\epsilon = -1.6 + 14.7i$ at a working frequency of $\omega = 2\pi \times 10^{14}$ Hz. I illuminate the sample with a monochromatic plane wave of that frequency and want to know which charges and currents induced in that sample are free and which are bound. According to the various definitions discussed above, there is no way to tell, unless more information is given about the microscopic origin of ϵ. But this additional microscopic information is neither required nor available in the macroscopic theory. Since this information is not needed, one can argue that the whole procedure of discriminating the free and bound charges is also not needed, since it does not lead to any observable physical consequences. This terminology is simply a legacy of 19th century physics.

However, there is a useful and, in fact, unavoidable distinction between two different types of charges and currents. We will call them external and induced. This will be discussed in detail in Sections 4–6.

3. MICROSCOPIC MAXWELL'S EQUATIONS AND LAWS OF MOTION

It is useful to start from the first principles, which, for our purposes, is the derivation of macroscopic Maxwell's equations from their microscopic counterparts. The microscopic equations are not subject of any controversy or ambiguity, and are of the form

$$\nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{J}, \qquad \nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}. \tag{1}$$

Here $\mathbf{E} = \mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B} = \mathbf{B}(\mathbf{r}, t)$ are, respectively, the electric and magnetic fields, which manifest themselves through the action of the Lorentz force and are, therefore, physically measurable quantities [26]. Consider N point charges $q_n, n = 1, 2, \ldots, N$. The Lorentz force acting on the nth point charge moving along the classical trajectory $\mathbf{r}_n(t)$ is

$$\mathbf{F}_n(t) = q_n\mathbf{E}(\mathbf{r}_n(t), t) + \frac{q_n}{c}\dot{\mathbf{r}}_n(t) \times \mathbf{B}(\mathbf{r}_n(t), t). \tag{2}$$

Here $\dot{\mathbf{r}}_n(t) = d\mathbf{r}_n(t)/dt = \mathbf{v}_n(t)$ is the velocity of the nth particle. The microscopic electric charge and current densities, $\rho = \rho(\mathbf{r}, t)$ and $\mathbf{J} = \mathbf{J}(\mathbf{r}, t)$ in Eq. (1), are defined, respectively, as

$$\rho(\mathbf{r}, t) = \sum_{n=1}^{N} q_n\delta(\mathbf{r} - \mathbf{r}_n(t)), \tag{3a}$$

$$\mathbf{J}(\mathbf{r}, t) = \sum_{n=1}^{N} q_n\dot{\mathbf{r}}_n(t)\delta(\mathbf{r} - \mathbf{r}_n(t)). \tag{3b}$$

Note that the continuity equation

$$\nabla \cdot \mathbf{J}(\mathbf{r}, t) + \frac{\partial \rho(\mathbf{r}, t)}{\partial t} = 0 \tag{4}$$

follows from Eq. (3) and the two equations $\nabla \cdot \mathbf{B} = 0$ and $\nabla \cdot \mathbf{E} = 4\pi\rho$, in turn, follow from Eqs. (1) and (4). We can, therefore, view the set of equations (1) and (3) as being fundamental.

We can also state the relativistic law of motion, viz.,

$$\frac{d\mathbf{p}_n(t)}{dt} = \mathbf{F}_n(t), \qquad \mathbf{p}_n(t) = \frac{m_n\mathbf{v}_n(t)}{\sqrt{1 - v_n^2(t)/c^2}}, \tag{5}$$

where m_n is the mass of the nth charge.

Thus, the self-consistent description of the classical (that is, relativistic but not quantum-mechanical) system of "point charges + fields" consists of the equations for the fields (1), the expression for the current (3b), the expression for the force acting on each charge (2), and the law of motion (5). Given appropriate initial conditions, these equations can be solved, at least, in principle, to describe uniquely the time evolution of the system. In particular, the trajectories $\mathbf{r}_n(t)$ and the fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathrm{r}, t)$ can be found in this manner. It can be shown that these solutions satisfy the conservation of the total energy E, the total linear momentum $\mathbf{P}$, and the total angular momentum $\mathbf{M}$. It can further be shown that the center of energy $\mathbf{R}(t)$ of the system (which replaces in the relativistic physics the Newtonian center of mass) satisfies the classical law of motion $E\dot{\mathbf{R}}(t) = c^2\mathbf{P}$. It is not our goal to derive these conservation laws here, but we note that they are mathematical consequences of Eqs. (1)–(5). Of course, the energy and momentum of the field as well as those of the particles must be accounted for to obtain the conservation laws. The point here is that, in contrast, the same conservation laws cannot be derived in the *macroscopic* theory without introducing some phenomenological terms. This observation indicates that the macroscopic theory cannot be derived from its microscopic counterpart in a completely straightforward mathematical manner and without making some phenomenological assumptions. This is discussed in more detail in Section 5.

4. EXTERNAL AND INDUCED CURRENTS

The physical model described in the previous section is self-consistent but is often not complete. In many physical situations, there exist some currents that cannot be included in the self-consistent dynamics of the system, but are assumed to be given and known, whether they are controlled by the experimentalist or not. We will call such currents and the fields generated by them *external* (to the system under consideration). All other currents and fields will be *induced*.

The external currents can be defined by the following two properties:

- In any particular electromagnetic problem, the external currents or the fields generated by the external currents (in the applicable domain) are known *a priori*, that is, before solving the problem.
- The external currents or the fields generated by the external currents are not influenced in any way by the solution to a given electromagnetic problem. In other words, the external currents are independent of the fields. As a consequence, the external currents do not depend on the type and physical properties of the system that we want to study.

Whether these currents are directly controlled by the experimentalist is not important. It is also not important whether these currents are conductivity currents, and, more generally, what the physical nature of the charges that create these currents is.

Examples of external currents include the current in the coils of a particle accelerator, or the current in a transmitting antenna, or the current flowing in the active medium of a laser. In Fig. 1, we illustrate the concept of external current using the example of a transmitting antenna. Here the current $\mathbf{J}_{\mathrm{ext}}$ is created, e.g., by connecting the antenna to a generator. However, we do not care and do not need to know how this current was created; all that is important for us is that $\mathbf{J}_{\mathrm{ext}}$ is known and is not influenced by the currents induced in any of the receiving antennas (there is no back-action). This condition can be satisfied in practice with extremely high precision. The receiving antennas are, however, considered self-consistently by solving Maxwell's equations. Therefore, the current in the receiving antennas is *induced* rather than external. We note that, if we place several receiving antennas in close proximity to each other, multiple-scattering effects will play an important role and the receiving antennas would no longer be isolated from each other.

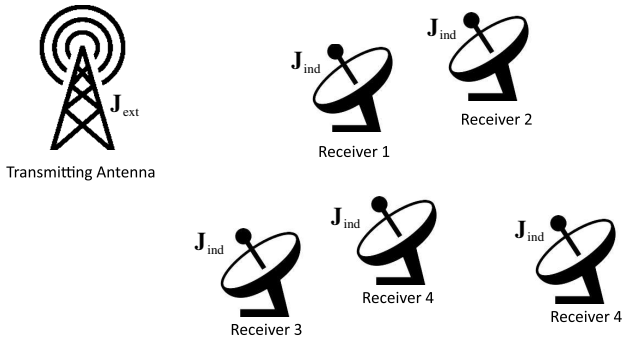

Fig. 1. Illustration of the concept of external current. The current in the transmitting antenna (external) is given and known and is not influenced by the receiving antennas. The currents in the receiving antennas (induced) can be computed by solving the appropriate boundary-value electromagnetic problem and depend on the geometry and physical properties of the receiving antennas.

It can be seen that the definition of the external current is not "physical" but, rather, situative. In some problems, the current in the transmitting antenna should be defined as external. In other problems, when we wish to consider the transmitting antenna itself in more detail, this current should be defined as induced. Then we still need to define another external current, perhaps, inside the generator. In fact, if we do not introduce *some* external current into the model, the solution to Maxwell's equations would be trivial: all currents and fields will be zero. The ultimate choice, which current is external and which is induced, depends on the problem we wish to solve and not on the physical nature of this current. Of course, the currents in all antennas are, primarily, the conductivity currents, and they satisfy Ohm's law $\mathbf{J} = \sigma\mathbf{E}$ locally. But in some problems this current is introduced in Maxwell's equations as external and in others as induced. For example, it is not possible or practical to consider the system "transmitting antenna + receiving antenna" self-consistently. Therefore, the current in the transmitting antenna is in this case external, and in the receiving antennas it is induced.

The external fields $\mathbf{E}_{\mathrm{ext}}$ and $\mathbf{B}_{\mathrm{ext}}$ are defined as solutions to the microscopic Maxwell's equations in which $\mathbf{J}_{\mathrm{ext}}$ serves as the source:

$$\nabla \times \mathbf{B}_{\mathrm{ext}} = \frac{1}{c}\frac{\partial \mathbf{E}_{\mathrm{ext}}}{\partial t} + \frac{4\pi}{c}\mathbf{J}_{\mathrm{ext}}, \quad \nabla \times \mathbf{E}_{\mathrm{ext}} = -\frac{1}{c}\frac{\partial \mathbf{B}_{\mathrm{ext}}}{\partial t}. \tag{6}$$

In many cases, the exact form of $\mathbf{J}_{\text{ext}}$ is not known, but the fields $\mathbf{E}_{\text{ext}}$ and $\mathbf{B}_{\text{ext}}$ are known. Moreover, it is often sufficient to know these fields in some bounded region of space. Note that $\mathbf{J}_{\text{ext}}(\mathbf{r},t)$ defines $\mathbf{E}_{\text{ext}}(\mathbf{r},t)$ and $\mathbf{B}_{\text{ext}}(\mathbf{r},t)$ uniquely everywhere in space. Conversely, $\mathbf{E}_{\text{ext}}(\mathbf{r},t)$ and $\mathbf{B}_{\text{ext}}(\mathbf{r},t)$ uniquely define $\mathbf{J}_{\text{ext}}(\mathbf{r},t)$, as follows from Eq. (6). However, if $\mathbf{E}_{\text{ext}}(\mathbf{r},t)$ and $\mathbf{B}_{\text{ext}}(\mathbf{r},t)$ are known only in some region of space, say, Ω, the problem of reconstructing $\mathbf{J}_{\text{ext}}(\mathbf{r},t)$ is mathematically unstable as any problem of analytical continuation. Thus, there can exist many different external current distributions that produce almost the same external fields (up to exponentially small variations) in Ω. In the scattering theory, the external fields are considered as known and given, whereas the external current is not considered directly. However, we should keep in mind that the external fields are always generated by the external current.

We now replace any of the receiving antennas in Fig. 1 with a microscopic system of point charges q_n and decompose the total current $\mathbf{J}$ in Eq. (1) into the external and induced components according to

$$\mathbf{J}(\mathbf{r},t)=\mathbf{J}_{\text{ext}}(\mathbf{r},t)+\mathbf{J}_{\text{ind}}(\mathbf{r},t). \tag{7}$$

Here $\mathbf{J}_{\text{ind}}$ is given by an expression identical to Eq. (3b), except that now the summation runs only over the charges belonging to the "receiving antenna" and does not include the charges that created the external current. Consequently, the sum yields only the induced current rather than the total current $\mathbf{J}$, viz.,

$$\mathbf{J}_{\text{ind}}(\mathbf{r},t)=\sum_{n\in\text{RA}} q_n\dot{\mathbf{r}}_n(t)\delta(\mathbf{r}-\mathbf{r}_n(t)). \tag{8}$$

The notation $n\in\text{RA}$ is a reminder that the summation runs only over the charges of the "receiving antenna" (a notion that we will generalize in what follows). The total fields are also decomposed into the external and induced components,

$$\mathbf{E}(\mathbf{r},t)=\mathbf{E}_{\text{ext}}(\mathbf{r},t)+\mathbf{E}_{\text{ind}}(\mathbf{r},t), \tag{9a}$$

$$\mathbf{B}(\mathbf{r},t)=\mathbf{B}_{\text{ext}}(\mathbf{r},t)+\mathbf{B}_{\text{ind}}(\mathbf{r},t), \tag{9b}$$

and the induced fields satisfy

$$\nabla\times\mathbf{B}_{\text{ind}}=\frac{1}{c}\frac{\partial\mathbf{E}_{\text{ind}}}{\partial t}+\frac{4\pi}{c}\mathbf{J}_{\text{ind}},\quad \nabla\times\mathbf{E}_{\text{ind}}=-\frac{1}{c}\frac{\partial\mathbf{B}_{\text{ind}}}{\partial t}. \tag{10}$$

The force acting on each charge, q_n, is now given by

$$\mathbf{F}_n(t)=q_n[\mathbf{E}_{\text{ext}}(\mathbf{r}_n(t),t)+\mathbf{E}_{\text{ind}}(\mathbf{r}_n(t),t)]+\frac{q_n}{c}\dot{\mathbf{r}}_n(t)[\mathbf{B}_{\text{ext}}(\mathbf{r}_n(t),t)+\mathbf{B}_{\text{ind}}(\mathbf{r}_n(t),t)]. \tag{11}$$

The laws of motion for the nth charge remain the same as before and are given by Eq. (5).

Let us now assume that $\mathbf{E}_{\text{ext}}(\mathbf{r},t)$ and $\mathbf{B}_{\text{ext}}(\mathbf{r},t)$ are known in a sufficiently large region of space Ω, which can be expected to contain all the trajectories $\mathbf{r}_n(t)$, at least, for some period of time. Then we have the following self-consistent formulation of the microscopic problem of "point charges moving in the external and self-induced fields." The equations for the induced fields are (10), where $\mathbf{J}_{\text{ind}}$ is given by (8); the force acting on each charge is given now by (11) and the equations of motion are still given by (5). Note that the conservation laws mentioned in Section 3 no longer hold because of the action of the external fields.

5. FROM MICROSCOPIC TO MACROSCOPIC MAXWELL'S EQUATIONS

It is widely believed that macroscopic Maxwell's equations can be derived by averaging their microscopic counterparts, e.g., Eq. (10). The averaging operation for a generic field $F(\mathbf{r})$ is usually defined as

$$\langle F(\mathbf{r})\rangle=\int S(\mathbf{r}')F(\mathbf{r}+\mathbf{r}')\mathrm{d}^3r', \tag{12}$$

where $S(\mathbf{r})$ is a mollifier with the properties $\int S(\mathbf{r})\mathrm{d}^3r=1$ and $S(\mathbf{r})\to 0$ when $r\to\infty$ (another possible approach is ensemble averaging, but it usually involves not the field itself but two-point and higher-order correlation functions [27]). In what follows, we will show that this belief is incorrect. Then the macroscopic equations will be derived from several phenomenological postulates, which are independent of the averaging argument.

We start by noting that, in any realistic material (except, perhaps, dilute gases), the microscopic fields that fluctuate on the atomic scale are not known and, therefore, their averaging is a purely formal procedure: it cannot be used, for example, to compute the permittivity of the material. DFT simulations can potentially be applied to this end, but DFT already operates with continuous densities of charge, at least, for the electrons. Still, one can argue that, even though the microscopic fields may not be known, their suitably defined averages can be shown to obey macroscopic Maxwell's equations. Indeed, it is true that the operation of spatial averaging (12) and the derivatives in Eq. (10) commute. However, the commutative property is not sufficient to show that macroscopic Maxwell's equations are averages of Eq. (10). It is also required that the microscopic fields be integrable so that the averages in question actually *exist*. But this is not the case. For example, it is easy to see that the electric field of a point charge is not integrable.

To illustrate the above point, consider an electrically neutral macroscopic sphere made of some spatially uniform dielectric material. In the absence of any external field, the sphere is not polarized and the macroscopic induced electric field is zero everywhere in space. Any correct averaging of the microscopic electric field must yield the same result. Obviously, the spatial integral of zero electric field over the entire space is also zero. We expect to obtain the same global zero result in the microscopic picture as well. Indeed, the exact and averaged functions must have the same global integrals, unless we have done something seriously wrong. Let us then integrate the electric field created by each point charge that makes up the sphere and add the results. But, as noted above, the electric field of a point charge is not integrable. We can attempt to regularize the integral by introducing small-distance and large-distance cut-offs, or by performing angular integration first and radial integration second, assuming the charge is at the origin of a reference frame. Whatever strategy we try, the only reasonable result of such a regularized integration is zero. This follows immediately from the isotropy of space. The sum of zeros is zero, and we see that the two global integrals in the macroscopic and microscopic pictures are consistent. Now, let us turn on some spatially uniform external electric field so that the sphere gets polarized and acquires a macroscopic dipole moment $\mathbf{d}_{\text{tot}}$, and then evaluate the two global integrals again. In the

microscopic picture, the point charges will shift from their equilibrium positions, but the regularized global integral for each charge will remain zero. However, in the macroscopic picture, the global integral of the induced electric field will now be equal to $-(4\pi/3)\mathbf{d}_{\text{tot}}$ (assuming we use a reference frame whose origin is the sphere's center and perform angular integration first). We thus have come to a contradiction.

It may not be immediately obvious what is going on in this example. There is, however, no paradox. The discrepancy has occurred because we manipulated diverging integrals and the manipulation was not mathematically rigorous. There is, in fact, no mathematically rigorous method to compute these integrals: they simply do not exist. Consequently, the macroscopic induced field *is not* the average of its microscopic counterpart because the latter does not exist.

Finally, consider the following point: even if the microscopic field could be spatially averaged with some mathematical trick, the result thus obtained would have no physical significance. Indeed, the expression for the Lorentz force contains the electric and magnetic fields at the exact locations of the charged particles, not some abstract averaged quantities. The particle trajectories and the microscopic electromagnetic fields are not statistically independent of each other. The mean-field approximation is inapplicable to the microscopic electrodynamics of realistic materials. Then why would anyone be interested in the averaged field? How can it be measured?

In fact, macroscopic Maxwell's equations can be derived from Eq. (10) by making an approximation to the induced current $\mathbf{J}_{\text{ind}}$. This is the only approximation that is involved; the fields $\mathbf{E}$ and $\mathbf{B}$ are computed from the resulting equations self-consistently. Obviously, these fields are also in some sense approximate, but they are not obtained by averaging of anything. The approximation for $\mathbf{J}_{\text{ind}}$ is based on the following three postulates (reproduced here with small clarifications from Ref. [25]):

1. The medium is a true continuum with piece-wise spatially uniform properties but, possibly, with sharp boundaries or interfaces where the properties can jump abruptly.
2. The external source excites a *continuous* density of induced current $\mathbf{J}_{\text{ind}}$ (possibly with singular contributions at the surfaces of discontinuity), which is a functional of the fundamental electromagnetic fields $\mathbf{E}$ and $\mathbf{B}$ (total fields, not just the induced components).
3. The induced current is zero in vacuum.

It can be pointed out that Postulate 1 is certainly an approximation. Discreteness of matter reveals itself in a variety of physical phenomena, which include Rayleigh scattering from random (incoherent) thermal fluctuations in gases and liquids or from frozen inhomogeneities in solid amorphous media, and x-ray diffraction in crystal.

Let the functional that maps the fundamental fields $\mathbf{E}$ and $\mathbf{B}$ be $\mathcal{F}[\cdot,\cdot]$, so that $\mathbf{J}_{\text{ind}}(\mathbf{r},t)=\mathcal{F}[\mathbf{E}(\mathbf{r},t),\mathbf{B}(\mathbf{r},t)]$. The question then arises how to find $\mathcal{F}[\cdot,\cdot]$. This involves some additional phenomenological approximations such as linearity, locality, and causality. These requirements can also be considered as postulates. For example, there is no *a priori* reason to believe that the electric field is the cause and the induced current is the effect and not *vice versa*. However, very extensive experimental evidence suggests that these postulates are quite accurate, and investigation of the cases wherein they break down leads to separate fields of research such as nonlinear optics or the theory of nonlocality. In the forthcoming discussion, we will consider the above postulates or approximations more carefully.

Linearity implies that $\mathcal{F}[\mathbf{E}_1+\mathbf{E}_2,\mathbf{B}_1+\mathbf{B}_2]=\mathcal{F}[\mathbf{E}_1,\mathbf{B}_1]+\mathcal{F}[\mathbf{E}_2,\mathbf{B}_2]$. Locality is traditionally interpreted as the requirement that $\mathcal{F}[\cdot,\cdot]$ contain no spatial integrals and no more than the first-order spatial derivative. Causality implies that $\mathbf{J}_{\text{ind}}(\mathbf{r},t)$ is not influenced in any way by $\mathbf{E}(\mathbf{r},t')$ and $\mathbf{B}(\mathbf{r},t')$, where $t'>t$. An additional requirement follows from symmetry: $\mathbf{J}_{\text{ind}}$ is a true vector (unlike the pseudo-vector $\mathbf{B}$) so that its Cartesian components must change sign under the operation of coordinate inversion. For simplicity, we add the requirements that the medium is *reciprocal*, *non-chiral*, and *non-gyrotropic* (actually, magnetic gyrotropy contradicts the requirement of linearity). Here we do not discuss possible exceptions to these requirements as this will lead us to the vast topic of bi-isotropic and bi-anisotropic media. However, in many natural materials, all these conditions hold with good precision. Then the most general form of the functional dependence of the induced current on the fundamental fields, which is consistent with the several assumptions made above, is of the form

$$\mathbf{J}_{\text{ind}}(\mathbf{r},t)=\frac{\partial\mathbf{P}(\mathbf{r},t)}{\partial t}+c\nabla\times\mathbf{M}(\mathbf{r},t), \tag{13}$$

where

$$\mathbf{P}(\mathbf{r},t)=\int_0^\infty \hat{f}_e(\mathbf{r},\tau)\mathbf{E}(\mathbf{r},t-\tau)\mathrm{d}\tau, \tag{14a}$$

$$\mathbf{M}(\mathbf{r},t)=\int_0^\infty \hat{f}_m(\mathbf{r},\tau)\mathbf{B}(\mathbf{r},t-\tau)\mathrm{d}\tau. \tag{14b}$$

Here the *influence functions* $\hat{f}_e(\mathbf{r},\tau)$ and $\hat{f}_m(\mathbf{r},\tau)$ are symmetric tensors. In isotropic media, the influence functions are reduced to scalars so that $\hat{f}_{e,m}=f_{e,m}\hat{I}$, where $\hat{I}$ is the identity tensor. The influence functions characterize the macroscopic response of a medium completely. Collectively, Eqs. (13) and (14) are known as the constitutive relations.

According to Postulate 1, the functions $\hat{f}_{e,m}(\mathbf{r},\tau)$ are spatially uniform inside a homogeneous material, but can jump abruptly at boundaries or interfaces. According to Postulate 2, $\hat{f}_e(\mathbf{r},\tau)$ is differentiable with respect to τ for all $\tau>0$, whereas $\hat{f}_m(\mathbf{r},\tau)$ is differentiable with respect to $\mathbf{r}$ everywhere except at boundaries and interfaces. According to Postulate 3, $\hat{f}_e(\mathbf{r},\tau)=\hat{f}_m(\mathbf{r},\tau)=\mathbf{P}(\mathbf{r},t)=\mathbf{M}(\mathbf{r},t)=0$ if $\mathbf{r}$ is in vacuum. In the most typical case of piece-wise homogeneous materials, $\hat{f}_{e,m}(\mathbf{r},\tau)$ are also piece-wise constant as functions of $\mathbf{r}$ and zero in vacuum. The more exotic case of media whose properties change smoothly in space can also be accounted for by the constitutive relations (13) and (14).

Sometimes the term $\sigma\mathbf{E}(\mathbf{r},t)$ is added to the right-hand side in Eq. (13), but we show in Section 7 that it can be absorbed in $\partial\mathbf{P}(\mathbf{r},t)/\partial t$ without loss of generality even in the static limit.

To derive macroscopic Maxwell's equations, we add Eqs. (6) and (10) together and substitute expression (13) for $\mathbf{J}_{\text{ind}}$.

This yields

$$\nabla \times (\mathbf{B} - 4\pi\mathbf{M}) = \frac{1}{c}\frac{\partial(\mathbf{E} + 4\pi\mathbf{P})}{\partial t} + \frac{4\pi}{c}\mathbf{J}_{\text{ext}}, \tag{15a}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}. \tag{15b}$$

We then define the auxiliary fields $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$ and $\mathbf{H} = \mathbf{B} - 4\pi\mathbf{M}$, and arrive at the conventional form of macroscopic Maxwell's equations:

$$\nabla \times \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} + \frac{4\pi}{c}\mathbf{J}_{\text{ext}}, \qquad \nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}. \tag{16}$$

The permittivity and permeability tensors can be defined in the frequency domain. Let all fields be Fourier-transformable with respect to time according to the convention

$$\tilde{\mathbf{E}}(\mathbf{r},\omega) = \int_{-\infty}^{\infty} \mathbf{E}(\mathbf{r},t)e^{i\omega t}\mathrm{d}t, \quad \mathbf{E}(\mathbf{r},t) = \int_{-\infty}^{\infty} \tilde{\mathbf{E}}(\mathbf{r},\omega)e^{-i\omega t}\frac{\mathrm{d}\omega}{2\pi},$$

where $\tilde{\mathbf{E}}(\mathbf{r}, -\omega) = \tilde{\mathbf{E}}^*(\mathbf{r},\omega)$, and similarly for other fields. Then Eq. (16) takes the form

$$\nabla \times \tilde{\mathbf{H}} = -\frac{\omega}{c}\tilde{\mathbf{D}} + \frac{4\pi}{c}\tilde{\mathbf{J}}_{\text{ext}}, \qquad \nabla \times \tilde{\mathbf{E}} = \frac{\omega}{c}\tilde{\mathbf{B}}, \tag{17}$$

where

$$\tilde{\mathbf{D}}(\mathbf{r},\omega) = \hat{\epsilon}(\mathbf{r},\omega)\tilde{\mathbf{D}}(\mathbf{r},\omega), \quad \tilde{\mathbf{H}}(\mathbf{r},\omega) = \hat{\mu}^{-1}(\mathbf{r},\omega)\tilde{\mathbf{B}}(\mathbf{r},\omega), \tag{18}$$

and the permittivity and permeability tensors $\hat{\epsilon}$ and $\hat{\mu}$ are expressed in terms of the influence functions as, respectively,

$$\hat{\epsilon}(\mathbf{r},\omega) = \hat{I} + 4\pi\int_0^{\infty} \hat{f}_e(\mathbf{r},\tau)\exp(i\omega\tau)\mathrm{d}\tau, \tag{19a}$$

$$\hat{\mu}(\mathbf{r},\omega) = \left[\hat{I} + 4\pi\int_0^{\infty} \hat{f}_m(\mathbf{r},\tau)\exp(i\omega\tau)\mathrm{d}\tau\right]^{-1}. \tag{19b}$$

A matrix inversion is implied in Eq. (19b). The same expressions can be obtained for truly monochromatic fields oscillating at the frequency of ω when Fourier transforms, strictly speaking, do not exist or exist in the sense of generalized functions.

This completes the derivation of macroscopic Maxwell's equations from their microscopic counterparts. One can argue that there is no real "derivation" involved. The only approximation or assumption that we have used was that the induced current $\mathbf{J}_{\text{ind}}$ in a continuous medium is expressible in terms of the fundamental fields $\mathbf{E}$ and $\mathbf{B}$ according to Eqs. (13) and (14), respectively.

One point that is frequently discussed in conjunction with expression (13) is its non-uniqueness or ambiguity [1,28]. By non-uniqueness, the following mathematical property is meant: let $\mathbf{P}' = \mathbf{P} + \nabla \times \mathbf{F}$ and $\mathbf{M}' = \mathbf{M} - \partial\mathbf{F}/\partial t$, where $\mathbf{F}$ is an arbitrary pseudo-vector field. Then we have $\partial\mathbf{P}/\partial t + c\nabla \times \mathbf{M} = \partial\mathbf{P}'/\partial t + c\nabla \times \mathbf{M}'$. Apparently, there are infinitely many pairs of fields $\{\mathbf{P},\mathbf{M}\}$, which, when substituted into Eq. (13), produce exactly the same induced current $\mathbf{J}_{\text{ind}}$. The question that arises then is how can we determine $\mathbf{P}$ and $\mathbf{M}$ uniquely for a given material and how does this non-uniqueness influence the determination of the medium parameters $\hat{\epsilon}$ and $\hat{\mu}$.

The answer is rather simple: there is never a need to determine $\mathbf{P}$ and $\mathbf{M}$ from a given $\mathbf{J}_{\text{ind}}$, and the macroscopic theory neither requires nor depends on the existence of a one-to-one correspondence between $\mathbf{J}_{\text{ind}}$ and the pair $\{\mathbf{P},\mathbf{M}\}$. Rather, in order for the theory to be consistent, one can expect the following property to hold: for every external excitation $\mathbf{J}_{\text{ext}}$, there is a unique solution to macroscopic Maxwell's equations, including the fundamental fields $\mathbf{E}$ and $\mathbf{B}$, and the auxiliary fields $\mathbf{D}$, $\mathbf{H}$, $\mathbf{P}$, and $\mathbf{M}$. This is indeed the case. It is also required that the influence functions for a given material be determined uniquely. In other words, two samples of the same shape and two different sets of parameters $\{\hat{f}_e,\hat{f}_m\}$ and $\{\hat{f}_e',\hat{f}_m'\}$ must be physically distinguishable unless $\{\hat{f}_e,\hat{f}_m\} = \{\hat{f}_e',\hat{f}_m'\}$. In fact, an even stronger statement can be made: at any fixed frequency ω, two samples of the same shape and two different sets of parameters $\{\hat{\epsilon}(\omega),\hat{\mu}(\omega)\}$ and $\{\hat{\epsilon}'(\omega),\hat{\mu}'(\omega)\}$ are physically distinguishable unless $\{\hat{\epsilon}(\omega),\hat{\mu}(\omega)\} = \{\hat{\epsilon}'(\omega),\hat{\mu}'(\omega)\}$. Therefore, the permittivity and permeability of a material are measurable quantities.

6. PROPERTIES OF THE EXTERNAL CURRENT AND SOME OF THEIR CONSEQUENCES

If one wishes to solve macroscopic Maxwell's equations for an arbitrary sample of continuous medium, then there must be some external current in the problem; otherwise, the solution is trivial. As discussed above, it is often sufficient to know the external field generated by this external current, not the current itself. However, the external current must exist as a matter of principle. Moreover, the external current cannot overlap spatially with the sample under consideration and cannot exchange any electric charge with it.

The above statement is important. Formally, it is possible to consider a mathematical model in which the external current in Eq. (16) or in its frequency-domain version (17) is arbitrary and, in particular, overlaps with the medium. This model can have a perfectly valid mathematical solution. However, this solution does not correspond to any physical system. The current inside any sample of continuous medium in which macroscopic Maxwell's equations are solved is by definition the induced current, and it is not arbitrary but governed by the constitutive relations. The latter define the medium's electromagnetic properties completely. If one allows some other current (not included in or not satisfying the constitutive relations), then the medium is effectively modified, it is no longer the same medium as the one described by the constitutive relations.

We can further conclude that the model of infinite unbounded medium can be introduced in the macroscopic theory only with adequate care. Indeed, the external current must be supported *somewhere*. Since it cannot be supported inside the medium, it must be supported outside of it. This excludes the possibility of infinite media in the classical electrodynamics. However, this model is deeply ingrained in the literature and it is important to understand when its use is justified.

We start by noting that the solution to macroscopic Maxwell's equations in some finite sample occupying the region Ω can be locally (that is, in $V \subset \Omega$) well approximated by a plane wave of the form $\mathrm{Re}[\mathbf{A}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}]$. On the other hand, a plane wave is a mathematical solution to homogeneous

Maxwell's equations (with $\mathbf{J}_{\text{ext}} = 0$) in an infinite medium. Therefore, one can forget about the boundaries of Ω and substitute the above plane-wave ansatz into macroscopic Maxwell's equations as if the medium is infinite. This will yield the dispersion relation and polarization modes of the medium, that is, the polarization vectors $\mathbf{A}_\alpha$ and the corresponding dispersion relations written implicitly as $f_\alpha(\mathbf{k}, \omega) = 0$ or explicitly as $\omega_\alpha = f_\alpha(\mathbf{k})$ or $\mathbf{k} = \mathbf{q}_\alpha(\omega)$, etc. Here α labels the plane-wave modes. The eigen-solutions thus obtained do not physically exist in the entire space, but can be good local approximations. Therefore, the model of infinite medium, as used above, is applicable and convenient, and for this reason it has gained wide acceptance.

Problems emerge when the model of infinite medium is used outside of its range of applicability, i.e., in problems that involve or depend on the medium impedance. Indeed, the impedance of an infinite medium is physically unobservable. For example, consider an isotropic infinite medium characterized by scalar permittivity and permeability of, respectively, $\epsilon(\omega)$ and $\mu(\omega)$. The dispersion relation for this medium can be obtained by using a plane-wave ansatz and is of the form $k^2 = (\omega/c)^2\epsilon(\omega)\mu(\omega)$. Note that this dispersion equation depends on the product of $\epsilon(\omega)$ and $\mu(\omega)$ but not on these functions individually. In other words, the impedance is unobservable if we exclude the medium boundaries from consideration.

We have discussed the condition of non-overlap of the external current and the material medium, and the limitations of the model of infinite media because mathematical models that transcend these limitations in various ways have been ubiquitous in the recent literature. It is useful, however, to trace some of the relevant ideas to the theory of electromagnetic nonlocality (spatial dispersion), which was developed in the 1960s, primarily, in the former USSR. For example, the proposition that an external current can spatially overlap with a sample of continuous medium in which macroscopic Maxwell's equations are being solved (and which, on top of that, is assumed to be infinite), can be traced to the well-known book *Crystal Optics with Spatial Dispersion and Excitons* [29] by Agranovich and Ginzburg (first published in 1965). On pages 24 and 25 of the cited English translation of the book, Agranovich and Ginzburg discuss the Fourier-space representation of the tensor $\hat{\epsilon}(\omega, \mathbf{k})$ in nonlocal materials, where ω and $\mathbf{k}$ are the Fourier variables reciprocal to t and $\mathbf{r}$, respectively, and it is worth adducing here the relevant text with only minor omissions (made for economy of space and displayed as […]):

> The variables ω and $\mathbf{k}$ on which the tensor $\hat{\epsilon}(\mathbf{k}, \omega)$ depends are, generally speaking, independent variables. This fact follows from the definition (∗), but sometimes this does not seem to be as clear as it might be. The point is that in optics we often deal with the propagation of waves in the absence of external sources in the medium itself. In this case, the wave vector depends on ω […] If, however, $\mathbf{k} = \mathbf{k}(\omega)$, spatial dispersion may seem to be equivalent to frequency dispersion. The answer to the question this poses is in the following. The tensor $\hat{\epsilon}(\mathbf{k}, \omega)$ is introduced for fields of the general form when the medium contains sources $\mathbf{J}_{\text{ext}}$ and ρ_{ext} […] Under these conditions a field $\mathbf{E}$ can be produced with any independent values of ω and $\mathbf{k}$.

$$\hat{\epsilon}(\mathbf{k}, \omega) = \int_0^\infty \mathrm{d}\tau \int \mathrm{d}^3 r e^{i(\mathbf{k}\cdot\mathbf{r} - \omega\tau)} \hat{\phi}_e(\mathbf{r}, \tau), \qquad (*)$$

$$\mathbf{D}(\mathbf{r}, t) = \int_{-\infty}^t \mathrm{d}t' \int \mathrm{d}^3 r' \hat{\phi}_e(\mathbf{r} - \mathbf{r}', t - t')\mathbf{E}(\mathbf{r}', t'). \qquad (**)$$

Here equations to which the quote refers have been written out using slightly different notations. Note that $\hat{\phi}_e(\mathbf{r}, \tau)$ is not the same function as $\hat{f}_e(\mathbf{r}, \tau)$ in Eq. (14a). In the case of $\hat{\phi}_e$, $\mathbf{r}$ denotes a shift of the coordinate, whereas in the case of $\hat{f}_e$, $\mathbf{r}$ is the position in space. One can say that the more general form of the influence function $\hat{f}_e$, which is applicable to the case of a spatially nonlocal medium, is $\hat{f}_e^{\text{nl}}(\mathbf{r}, \mathbf{r}', \tau)$. In a spatially uniform medium, when both arguments $\mathbf{r}$ and $\mathbf{r}'$ are sufficiently far from its boundaries, $\hat{f}_e^{\text{nl}}$ becomes a function of the shift $\mathbf{r} - \mathbf{r}'$ so that we can write $\hat{f}_e^{\text{nl}}(\mathbf{r}, \mathbf{r}', \tau) \approx \hat{\phi}_e(\mathbf{r} - \mathbf{r}', \tau)$. In the case of a strictly local medium, $\hat{f}_e^{\text{nl}}(\mathbf{r}, \mathbf{r}', \tau) = \hat{f}_e(\mathbf{r}, \tau)\delta(\mathbf{r} - \mathbf{r}')$.

The problem with the above quote is that it conflates two different statements, one of which is correct and the other is not. The correct statement is that $\hat{\epsilon}(\mathbf{k}, \omega)$ is a well-defined function of the mathematically independent arguments $\mathbf{k}$ and ω. Indeed, the relation (∗∗) is a valid generalization of the constitutive relations (14a) to the case of nonlocal media, as long as point $\mathbf{r}$ is sufficiently far from the boundaries of the media. It is also true that the function $\hat{\phi}_e(\mathbf{r}, t)$ can be Fourier-transformed according to (∗). The functions $\hat{\phi}_e(\mathbf{r}, \tau)$ and $\hat{\epsilon}(\mathbf{k}, \omega)$ are Fourier images of each other; whatever information is contained in one of them is also contained in the other. One can use $\hat{\epsilon}(\mathbf{k}, \omega)$ to find the dispersion relation. Obviously, one needs to know $\hat{\epsilon}(\mathbf{k}, \omega)$ for all values of its arguments for this purpose.

The quoted text becomes problematic when it suggests that, since $\hat{\epsilon}(\mathbf{k}, \omega)$ is a function of its two mathematically independent arguments, there must exist physical fields inside the medium that "probe" any particular combination of $\mathbf{k}$ and ω directly. It is further suggested that, otherwise, nonlocality is indistinguishable from frequency dispersion. This point is interesting and deserves discussion. Consider for simplicity an isotropic non-chiral medium. Then the dispersion relation takes the form

$$k^2 = (\omega/c)^2\epsilon_\perp(k, \omega), \qquad \textbf{(20)}$$

where the scalar function $\epsilon_\perp$ depends on the magnitude of $\mathbf{k}$ but not on its direction. Note that $\epsilon_\perp(k, \omega)$ is often referred to as the transverse permittivity. It describes the response of the medium to transverse plane waves. In the nonlocal theory, the wave number of a plane wave $\mathbf{k}$ defines a preferred direction in space. For this reason, the nonlocal permittivity $\hat{\epsilon}(\mathbf{k}, \omega)$ can be a tensor with two different "transverse" and "longitudinal" principal values even in a completely isotropic medium. The corresponding functions $\epsilon_\perp(k, \omega)$ and $\epsilon_\parallel(k, \omega)$ are, however, scalars and depend only on the magnitude of $\mathbf{k}$ but not on its direction. In the vast majority of natural materials, longitudinal waves are not supported and therefore $\epsilon_\parallel$ does not enter the dispersion relation. If the medium is not isotropic, then its description can become much more complicated. For example, the crystallographic axes can also serve as preferred directions in space (in monocrystalline solids). Different directions of propagation in such media are not equivalent. This form of nonlocality can result in dispersion relations and an isofrequency

surface that cannot be obtained in any purely local medium [30].

Returning to the discussion of dispersion relations in nonlocal media, the solution to Eq. (19) can be written as $k(\omega) = (\omega/c)\eta(\omega)$, where $\eta(\omega)$ is some function. In this case, nonlocality is indeed indistinguishable from frequency dispersion. In particular, the dispersion relation in the nonlocal medium characterized by some $\eta(\omega)$ is indistinguishable from that in a purely local medium with $\epsilon(\omega)\mu(\omega) = \eta(\omega)$. The authors of the quoted text above find this fact to be problematic for the theory. However, there is nothing wrong or surprising in it. The reason why nonlocality appears to be indistinguishable from frequency dispersion is that the discussion was limited so far to infinite media. As soon as one considers an actual problem with physical boundaries, it would become apparent that the solutions in the nonlocal sample characterized by some $\eta(\omega)$ and the local sample with $\epsilon(\omega)\mu(\omega) = \eta(\omega)$ are not the same. In fact, in the latter case the solutions will depend on $\epsilon(\omega)$ and $\mu(\omega)$ separately.

Returning to the question of external currents, we note that the theory of nonlocality does not depend in any way on the existence or feasibility of external currents overlapping with the medium. The nonlocal permittivity tensor $\hat{\epsilon}(\mathbf{k}, \omega)$ can be computed by Fourier transforming the influence function, and the latter can be obtained, for example, from some microscopic model, symmetry considerations, or phenomenology. A method for computing this tensor in photonic crystals that is completely independent of the external current and does not require its consideration or introduction, has been described in Refs. [19,30]. On the other hand, introducing the external current overlapping with the medium will result in solutions that can never be excited in the actual sample. These solutions are perfectly valid mathematically, but they are solutions to a model that does not describe the physical reality. Potentially, one can use these solutions to compute the tensor $\hat{\epsilon}(\mathbf{k}, \omega)$ for arbitrary values of the arguments. This is not a mistake. However, treating plane waves with arbitrary $\mathbf{k}$ and ω as physical solutions that can exist in the medium, and computing for them various physical observables (e.g., energy-related quantities that are quadratic in the fields), is, in fact, a mistake. In the author's opinion, it is better to avoid confusion and not introduce external currents that overlap with the medium at all, as this is not required in any reasonable calculations and since these currents do not exist in nature. This is especially important in the theory of homogenization, where the so-called "current-driven model" has gained popularity in recent years.

7. ZERO FREQUENCY LIMIT AND STEADY CURRENTS

The last question we wish to address is whether the term $\sigma\mathbf{E}$ must be explicitly included in the expression for the induced current [Eq. (13)] or can always be absorbed in the term $\partial\mathbf{P}/\partial t$.

First, let us consider the question in the frequency domain. We still use the Fourier convention (18). Applying this convention to Eq. (13), we obtain

$$\tilde{\mathbf{J}}_{\text{ind}}(\mathbf{r}, \omega) = -i\omega\tilde{\mathbf{P}}(\mathbf{r}, \omega) + c\nabla \times \tilde{\mathbf{M}}(\mathbf{r}, \omega). \tag{21}$$

Let the medium be non-magnetic so that $\tilde{\mathbf{M}} = 0$. Recalling the definitions of the permittivity [Eq. (19a)], we can also rewrite this expression in the form

$$\tilde{\mathbf{J}}_{\text{ind}}(\mathbf{r}, \omega) = -i\omega\frac{\hat{\epsilon}(\mathbf{r}, \omega) - 1}{4\pi}\tilde{\mathbf{E}}(\mathbf{r}, \omega). \tag{22}$$

Let $\hat{\epsilon}$ inside a homogeneous sample occupying the region Ω be given by the well-known Drude formula

$$\hat{\epsilon}(\mathbf{r}, \omega) = \hat{I}\left[1 - \frac{\omega_p^2}{\omega(\omega + i\gamma)}\right] \quad \text{for } \mathbf{r} \in \Omega. \tag{23}$$

Here $\hat{I}$ is the unit tensor, ω_p is the plasma frequency, and γ is the relaxation constant. Substituting Eq. (23) into Eq. (22) and taking the limit $\omega \to 0$, we find that

$$\tilde{\mathbf{J}}_{\text{ind}}(\mathbf{r}, 0) = \frac{\omega_p^2}{4\pi\gamma}\tilde{\mathbf{E}}(\mathbf{r}, 0). \tag{24}$$

We see that, in the case of the Drude permittivity, the expression $\partial\mathbf{P}/\partial t$ has a well-defined static limit of the form $\sigma\mathbf{E}$, where $\sigma = \omega_p^2/4\pi\gamma$, and this limit indeed has the form of Ohm's law.

The above result may seem counter-intuitive. One might think that, in the case of a steady current, nothing is changing inside the material through which the current is flowing, so that $\mathbf{P} = \text{const}$ and $\partial\mathbf{P}/\partial t = 0$. For this reason, the belief that the term $\sigma\mathbf{E}$ must be explicitly included in Eq. (13) is rather widespread. However, it is simply not true that a steady current implies that $\mathbf{P} = \text{const}$.

To illustrate the above point, consider the simple circuit shown in Fig. 2. Here the electric charge accumulated in two conducting spheres, $+Q(t)$ and $-Q(t)$, is slowly discharging through a resistive wire. The distance between the charged spheres, $\Delta(t)$, can be changed by some external force of non-electromagnetic nature. We will initially consider the case when $\Delta(t)$ is fixed and constant, and then relax this condition. For simplicity, we also assume that the wire is uniform, that is, it is made of the same material and has a constant cross section. If the wire's resistance is sufficiently large, then the current flowing through the circuit would change in time very slowly and, in some time interval, it can be considered as constant.

Let us now assume that the current density inside the wire can be written in the form $\mathbf{J}_{\text{ind}} = \partial\mathbf{P}/\partial t$. We will show that this does not result in any contradictions. If the above assumption is true, then the induced charge can be computed from the continuity equation and is given by $\rho_{\text{ind}} = -\nabla \cdot \mathbf{P}$. In particular, the charge in the spheres is induced: it was accumulated due

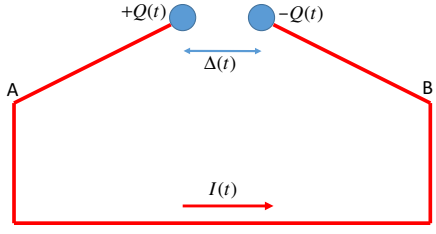

Fig. 2. Electric charge stored in two conducting spheres is slowly discharging through a resistive wire. The two wires connected to the charged spheres can swing in the plane of the drawing about the points A and B, thus changing the distance between the spheres, $\Delta(t)$, arbitrarily.

to the action of some external current or field in the distant past and is now slowly discharging through the system in complete agreement with macroscopic Maxwell's equations. Then the total dipole moment of the system is

$$\mathbf{d}_{\text{tot}}(t) = \int_V \mathbf{P}(\mathbf{r}, t)\mathrm{d}^3 r, \tag{25}$$

where V is the spatial region occupied by the wire and spheres. This is a well-known result; it is valid for all objects that are electrically neutral as a whole. Therefore, the vector of polarization in the system is nonzero. Let the polarization of the wire be given by the vector $\mathbf{P}(\ell, t)$ of ℓ-independent amplitude $P(t)$, where ℓ is the coordinate along the wire, with $\ell = 0$ being the coordinate just at the point where the wire is attached to the negatively charged sphere and $\ell = L$ being the coordinate at the point where the wire is attached to the positively charged sphere, and L is the total length of the wire. Obviously, the vector of polarization is always oriented along the wire and points in the direction from the negative charge toward the positive charge. Therefore, the amplitude of polarization is constant along the wire at any given moment of time, but the direction can change. Then it is not difficult to find that

$$\mathbf{d}_{\text{tot}}(t) = P(t)\Delta(t)S. \tag{26}$$

Here S is the wire's cross-section area, and this result applies to any geometry of the wire, even if it can bend out of plane.

On the other hand, the dipole moment of the system can be computed geometrically and is equal to

$$\mathbf{d}_{\text{tot}}(t) = Q(t)\Delta(t). \tag{27}$$

Comparing Eqs. (26) and (27), we find that

$$P(t) = Q(t)/S. \tag{28}$$

Thus, if the charge $Q(t)$ changes in time, so does the amplitude of polarization of the wire, $P(t)$.

Let us first consider the case when the distance between the charged spheres is fixed so that $\Delta(t) = \Delta_0$. Then the current flowing through the wire is $I(t) = \dot{Q}(t)$, where the prime denotes differentiation with respect to time. On the other hand, the current density in the wire is $\mathbf{J}_{\text{ind}}(\ell, t) = \dot{\mathbf{P}}(\ell, t)$. The total current through any cross section of the wire is given by $I(t) = SJ_{\text{ind}}(t)$, where $J_{\text{ind}}(t)$ is the amplitude of the induced current density—just like the amplitude of $\mathbf{P}(\ell, t)$, it is ℓ-independent. We thus find that

$$I(t) = SJ_{\text{ind}}(t) = S\dot{P}(t) = S[\dot{Q}(t)/S] = \dot{Q}(t), \tag{29}$$

which is in full agreement with the former result. Here we have used Eq. (28) to replace $\dot{P}(t)$ with $\dot{Q}(t)/S$.

We thus see that the polarization in the wire actually changes at the same rate as the charge accumulated in the spheres. Now, if the dependence $Q(t)$ is linear in some time interval, e.g., $Q(t) \approx Q_0$-at for $t_1 < t < t_2$, then the current in that interval is constant. We therefore have an example where the physical state of a given piece of wire does not change at all, yet its polarization is changing. How is it possible?

The answer is that the vector of polarization is an auxiliary quantity and it is not defined locally by the physical state of the material. In particular, $\mathbf{P}$ *is not* the differential density of dipole moment (if that quantity could even be defined unambiguously). However, the integral relation (26) between the *total* dipole moment of an object and its vector of polarization holds.

One can argue that the current will never be truly constant in the above scheme because the voltage between the charged spheres will continuously drop. While this is not a strong argument (theoretically, the current can change as slowly as one wishes, although there might be some practical limitations), we can use some external non-electromagnetic force to change the distance between the charged spheres, $\Delta(t)$, in such a way that the voltage drop between the two spheres remains constant. Obviously, the distance must be increased to this end and some positive mechanical work must be spent. This is, in fact, the simplest possible realization of a generator: it uses some outside mechanical work converted to electromagnetic energy that keeps the current flowing through the wire strictly constant. Note that allowing $\Delta(t)$ to change slowly (so that no noticeable magnetic field is generated) does not influence any of the above arguments.

We finally note that the function $\hat{\epsilon}(\mathbf{r}, \omega)$ has a simple pole at $\omega = 0$ and for all values of $\mathbf{r}$ for which the medium is conducting. The residue at that pole defines the conductivity tensor at $\mathbf{r}$, $\hat{\sigma}(\mathbf{r})$ (up to the numerical factor of 4π). In the time domain, the influence function of a conducting medium has a contribution of the form $\hat{f}_\epsilon(\mathbf{r}, \tau) = \hat{\sigma}(\mathbf{r})\Theta(\tau)$, where $\Theta(x)$ is the unit step function.

8. SUMMARY

In the above sections, we have considered several fundamental questions of classical electrodynamics. It might be convenient to summarize some of the key points here.

First, the distinction between free and bound charges (currents) is the legacy of old physics and is not needed by the theory. The distinction can be used without making any substantive errors, but potential for such errors exists, especially when the concepts are used by beginning physicists. The utilization of these concepts often results in formulas that are lengthier or look more complicated than what is necessary. A lack of an unambiguous definition (which we argue cannot be given) complicates things further. The author argues that the notions of free and bound charges can be safely abandoned.

Second, the distinction between external and induced charges (currents) is much more fundamental and is, in fact, unavoidable. However, the definition of external currents is situative rather than physical. The same current (say, in a transmitting antenna) can be viewed as external in one problem but induced in another. Students should develop a solid physical intuition in this regard.

Third, macroscopic Maxwell's equations can be derived from several simple phenomenological postulates, which do not involve any field averaging. In fact, the field of any collection of point charges cannot be averaged due to the divergence of the relevant integrals.

Fourth, as follows from the definition of external current, it cannot spatially overlap with a material medium in which macroscopic Maxwell's equations are considered or solved.

Fifth, as follows from the previous statement, the model of infinite unbounded medium has serious limitations and does not allow a complete description of electromagnetic

phenomena occurring in continuous media. Consideration of boundaries is important and in many instances unavoidable. One particular example of a problem in which the consideration of boundaries is crucial is the so-called problem of homogenization.

Finally, there is no physical or mathematical need to consider the conductivity current $\sigma\mathbf{E}$ separately as a "free current." In some problems the conductivity current is external, whereas in other problems it is induced. If the latter is the case, then the conductivity current can be absorbed in the term $\partial\mathbf{P}/\partial t$ even in the static limit.

REFERENCES AND NOTES

1. A. Chipouline, C. Simovski, and S. Tretyakov, "Basics of averaging of the Maxwell equations for bulk materials," Metamaterials **6**, 77–120 (2012).
2. A. P. Vinogradov and A. M. Merzlikin, "Comment on "Basics of averaging of the Maxwell equations for bulk materials"," Metamaterials **6**, 121–125 (2012).
3. M. Anelli, D. Jonsson, H. Fliegl, and K. Ruud, "The origin dependence of the material constants: the permittivity and the inverse permeability," Molec. Phys. **113**, 1899–1913 (2015).
4. A. D. Yaghjian, "Classical power and energy relations for macroscopic dipolar continua derived from the microscopic Maxwell equations," Prog. Electromagn. Res. B **71**, 1–37 (2016).
5. M. Mansuripur, "Force, torque, linear momentum, and angular momentum in classical electrodynamics," Appl. Phys. A **123**, 653 (2017).
6. H. O. Hagenvik, K. Blotekjaer, and J. Skaar, "Dielectric media considered as vacuum sources," Am. J. Phys. **85**, 830–833 (2017).
7. M. G. Silveirinha, "Metamaterial homogenization approach with application to the characterization of microstructured composites with negative parameters," Phys. Rev. B **75**, 115104 (2007).
8. M. G. Silveirinha, "Poynting vector, heating rate, and stored energy in structured materials: a first principles derivation," Phys. Rev. B **80**, 235120 (2009).
9. J. T. Costa, M. G. Silveirinha, and S. I. Maslovski, "Finite-difference frequency-domain method for the extraction of effective parameters of metamaterials," Phys. Rev. B **80**, 235124 (2009).
10. C. Fietz and G. Shvets, "Current-driven metamaterial homogenization," Physica B **405**, 2930–2934 (2010).
11. C. Fietz and G. Shvets, "Homogenization theory for simple metamaterial as one-dimensional arrays of thin polarizable sheets," Phys. Rev. B **82**, 205128 (2010).
12. A. Alu, "First-principles homogenization theory for periodic metamaterials," Phys. Rev. B **84**, 075153 (2011).
13. J. T. Costa, M. G. Silveirinha, and A. Alu, "Poynting vector in negative-index metamaterials," Phys. Rev. B **83**, 165120 (2011).
14. M. G. Silveirinha, "Time domain homogenization of metamaterials," Phys. Rev. B **83**, 165104 (2011).
15. A. V. Chebykin, A. A. Orlov, A. V. Vozianova, S. I. Maslovski, Y. S. Kivshar, and P. A. Belov, "Nonlocal effective medium model for multilayered metal-dielectric metamaterials," Phys. Rev. B **84**, 115438 (2011).
16. C. Fietz and C. M. Soukoulis, "Scattering matrix of the boundary of a nonlocal metamaterial," Phys. Rev. B **86**, 085146 (2012).
17. A. A. Orlov, S. V. Zhukovsky, I. V. Iorsh, and P. A. Belov, "Controlling light with plasmonic multilayers," Photon. Nanostr. Fundam. Appl. **12**, 213–230 (2014).
18. V. A. Markel, "On the current-driven model in the classical electrodynamics of continuous media," J. Phys. Condens. Matter **22**, 485401 (2010).
19. V. A. Markel and I. Tsukerman, "Current-driven homogenization and effective medium parameters for finite samples," Phys. Rev. B **88**, 125131 (2013).
20. A. M. Merzlikin and R. S. Puzko, "Homogenization of Maxwell's equations in 1D beyond static approximation: self-averaging of the effective refractive index and Anderson localization of light," arXiv:1708.01661 (2017).
21. I. E. Tamm, *Fundamentals of the Theory of Electricity* (MIR, 1979).
22. E. M. Purcell, *Electricity and Magnetism*, 2nd ed. (McGraw-Hill, 1985).
23. R. Resta and D. Vanderbilt, "Theory of polarization: a modern approach," in *Physics of Ferroelectrics: a Modern Perspective* (Springer, 2007), pp. 31–68.
24. N. A. Spaldin, "A beginner guide to the modern theory of polarization," J. Sol. St. Chem. **195**, 2–10 (2012).
25. V. A. Markel, "Maxwell Garnett approximation (advanced topics): tutorial," J. Opt. Soc. Am. A **33**, 2237–2255 (2016).
26. Here we do not discuss the Aharonov–Bohm effect and other purely quantum phenomena.
27. L. A. Apresyan and Y. A. Kravtsov, *Radiation Transfer, Statistical and Wave Aspects* (Gordon and Breach, 1996).
28. L. P. Pitaevskii, "On analytical properties of the diamagnetic permeability in the presence of the spatial dispersion," Int. J. Quantum Chem. **112**, 2998–3001 (2012).
29. V. Agranovich and V. Ginzburg, *Spatial Dispersion in Crystal Optics, and Excitons* (Springer, 1984).
30. V. A. Markel and I. Tsukerman, "Applicability of effective medium description to photonic crystals in higher bands: theory and numerical analysis," Phys. Rev. B **93**, 224202 (2016).