Classical electrodynamics and the quantum nature of light

1997 J. Phys. A: Math. Gen. 30 6565

(http://iopscience.iop.org/0305-4470/30/18/030)

J. Phys. A: Math. Gen. **30** (1997) 6565–6585. Printed in the UK PII: S0305-4470(97)83742-0

Classical electrodynamics and the quantum nature of light

Manoelito M de Souza
Universidade Federal do Espírito Santo, Departamento de Física, 29065.900, Vitória, ES, Brasil

Received 28 April 1997

Abstract. A review of old inconsistencies of classical electrodynamics (CED) and some new ideas that solve them is presented. Problems with causality violating solutions of the wave equation, the electron equation of motion, and problems with the non-integrable singularity of its self-field energy tensor are well known. The correct interpretation of the two (advanced and retarded) Lienard–Wiechert solutions are in terms of the creation and annihilation of particles in classical physics. They are both retarded solutions. Previous work on the short-distance limit of CED of a spinless point electron are based on a faulty assumption which causes the well known inconsistencies of the theory: a diverging self-energy (the non-integrable singularity of its self-field energy tensor) and a causality-violating third-order equation of motion (the Lorentz–Dirac equation). The correct assumption fixes these problems without any change in the Maxwell's equations and let exposed, in the zero-distance limit, the discrete nature of light: the flux of energy from a point charge is discrete in time. CED cannot have a true equation of motion, only an effective one, as a consequence of the intrinsic meaning of the Faraday–Maxwell concept of field that does not correspond to the classical description of photon exchange, but only to the smearing of its effects in the space around the charge. This, in varied degrees, is transferred to QED and other field theories that are based on the same concept of fields as space-smeared interactions.

1. Introduction

Classical and quantum physics are considered to involve sharply distinct concepts and kinds of theories. Classical physics represents an approximation of the more refined and closer to a true description of the world which is supposedly done by quantum physics. The short-distance limit of both has always been plagued by unsurmountable problems, which in classical electrodynamics (CED) are attributed to the assumed point-like nature of the electron. Assuming a finite non-zero dimension for the electron brings, however, more problems than solutions. The blame is not on the point-like electron but on an incorrect approach of taking the theory zero-distance limit. A more careful approach will free the theory of these problems and will reveal some quantum aspects which until now were unsuspected in a classical theory.

The CED of a point electron is based on the Lienard–Wiechert solution (LWS); its many old and unsolved problems [1–3] make it a non-consistent theory. The Lienard–Wiechert advanced solution represents itself as a causality problem that has required some, at least verbal, efforts to be circumvented. One must also mention the field singularity or the self-energy problem; the non-integrable singularities of its energy tensor; the causality-violating behaviour of solutions of the Lorentz–Dirac equation [4–8]; etc. It will be shown here that the solution to these problems is connected to a more strict implementation of causality (extended causality), already present, although not yet recognized, in the LWS.

0305-4470/97/186565+21$19.50

In section 2 the notation is defined in a brief review of the standard interpretation of the two (advanced and retarded) LWSs. Causality can be seen as a restriction to access to regions of the spacetime manifold, as discussed in section 3, where the notion of extended causality is introduced; it allows a new interpretation of the two LWSs in terms of creation and annihilation of classical particles. The notion of a classical photon is introduced. In section 4 the singularities and the non-integrability of the electron self-field energy tensor as they are described in the literature are reviewed and discussed. It is then pointed that they are all consequences of using an implicit assumption about the zero-distance limit that will be proved to be faulty in the following section. Section 5 shows how to correctly take the zero-distance limit in CED and to give it a consistent physical interpretation. The anticipated recognition in the classical theory of the actual quantum nature of the electromagnetic radiation is necessary for having a clear physical picture behind these new mathematical results. Some algorithms, that will be used in the rest of the paper for taking the zero-distance limit, are presented in section 6. In section 7, while searching for an electron 'equation of motion', it is confirmed by an explicit direct calculation that the old problem of singularity and non-integrability of the electron self-field energy tensor has vanished just with correctly taking the zero-distance limit. In this paper we never tamper with the Maxwell's equations and the energy tensor. All that is allowed is a possible reinterpretation of their physical meaning. The most remarkable new feature is that the energy flux from a point charge is discrete in time, which requires an interpretation of light in terms of discrete emission of point-like objects (classical photons) and a revision of the physical meaning of the Gauss's law and Faraday–Maxwell concept of field. This will be done in the last section. The electron 'equation of motion', derived in section 8, does not have the problematic Schott term, it is just an effective equation. This is a consequence of the bilocal character of the LWS, as it depends on two points possibly far apart: the point where the signal is defined, and the point, in the source worldline, where it was created. It is then argued that CED cannot produce a true equation of motion for its sources as far as its formulation is based on the Faraday–Maxwell concept of fields. Section 9 is included as an appendix of section 8 to show an alternative calculation that enlightens its physical meaning. The paper is concluded in section 10 with a summary and a discussion of the physical content of the Maxwell–Faraday concept of field upon which the modern field theory is entirely based. Gauss's law is not compatible with the vision of a classical field in terms of exchange of discrete objects (classical photons) unless the fields represent rather space average effects taken over a period of time larger than the time interval among the photon emissions. The Faraday–Maxwell concept of field, which is based on the validity of the Gauss's law, represents the smearing over the space surrounding the charge of the effects of the exchanged photons. The field singularity at the charge position is a reflection of this smearing process or of the field space-average character.

2. The Lienard–Wiechert solutions

The retarded Lienard–Wiechert potential

$$A(x) = \frac{V}{\rho}\Big|_{\tau_{\text{ret}}} \qquad \text{for } \rho > 0 \tag{1}$$

is a (the retarded one) solution to the wave equation

$$\Box\, A(x) = 4\pi J(x) \tag{2}$$

and to the gauge condition,

$$\partial A \equiv \frac{\partial A^{\mu}}{\partial x^{\mu}} = 0 \tag{3}$$

where J, given by

$$J(x) = \int \mathrm{d}\tau\, V\delta^4[x - z(\tau)] \tag{4}$$

is the current for a point electron that describes a given trajectory $z(\tau)$, parametrized by its proper time τ; $V = \mathrm{d}z/\mathrm{d}\tau$. The electron charge and the speed of light are taken as 1.

$$\rho := -V_{\alpha}R^{\alpha} = -V\eta R = -VR \tag{5}$$

where η is the Minkowski metric tensor diag$(-1, 1, 1, 1)$, and $R := x - z(\tau)$. ρ is the invariant distance (in the charge rest frame) between $z(\tau_{\mathrm{ret}})$, the position of the charge at the retarded time, and x, its self-field event (see figure 1). The constraints

$$R^2 = 0 \tag{6}$$

and

$$R^0 > 0 \tag{7}$$

must be satisfied. The constraint $R^2 = 0$ requires that x and $z(\tau)$ belong to a same light-cone; it has two solutions, τ_{ret} and τ_{adv}, which are, respectively, the points where J intercepts the past and the future light-cone of x (see figure 1). The retarded solution describes a signal emitted at $z(\tau_{\mathrm{ret}})$ and that is being observed at x, with $x^0 > z^0(\tau_{\mathrm{ret}})$, while the advanced solution also observed at x, *will be* emitted in the future, at $z(\tau_{\mathrm{adv}})$, with $x^0 > z^0(\tau_{\mathrm{adv}})$. $R^0 > 0$ is a restriction to the retarded solution (1) as it excludes the causality violating advanced solution, and justifies the restriction $|_{\tau_{\mathrm{ret}}}$ in (1). But this is not the only available interpretation; it will be shown below another one that does not have problems with causality violation and that, remarkably, allows the description of particle creation and annihilation still in a classical physics context.

3. Causality and spacetime geometry

There is a well known geometric and physical interpretation of the constraint (6). $R^2 = 0$ assures that $A(x)$ is a signal that propagates with the speed of light, on a light-cone; in field theory it corresponds to the implementation of the so-called *local causality*: only points inside or on a same light-cone can be causally connected. For a physical object it defines, at a point, *its physical spacetime*, that is the regions of the space-time manifold that it can have access to. In the literature, only (6) is clearly associated to the notion of causality but this is not enough because $A(x)$, in CED, is just an ancillary intermediary step to the Maxwell stress tensor, to whose components, the electric and magnetic fields, are attributed the physical meaning of force carriers. So, it is necessary to consider variations of (1) and, therefore of (6). The constraint (6) must be considered in the neighbourhoods of x and z: $x + \mathrm{d}x$ and $z(\tau_{\mathrm{ret}} + \mathrm{d}\tau)$ must also belong to a same light-cone. Differentiation of (6) ($R\,\mathrm{d}R = 0 \rightarrow R(\mathrm{d}x - V\,\mathrm{d}\tau) = 0 \rightarrow R\,\mathrm{d}x + \rho\,\mathrm{d}\tau = 0$) generates the constraint

$$\mathrm{d}\tau + K\,\mathrm{d}x = 0 \tag{8}$$

where K, defined for $\rho > 0$ by

$$K := \frac{R}{\rho} \tag{9}$$

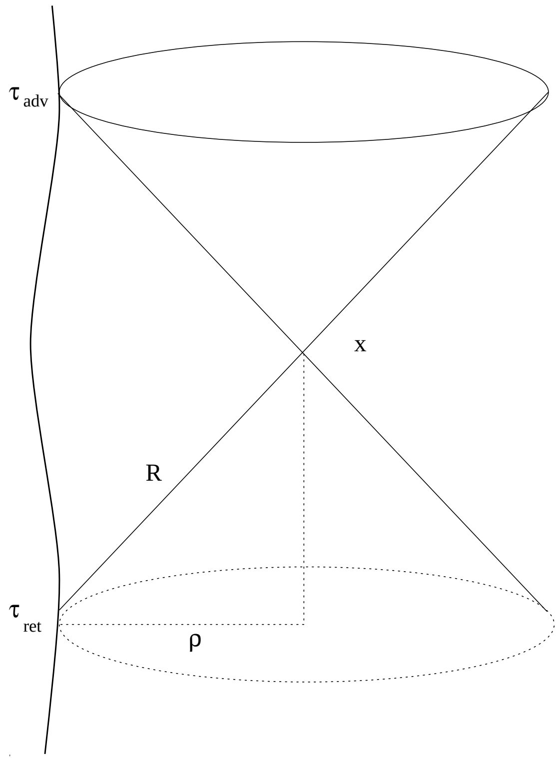

Figure 1. The usual interpretation of the LWSs. Two spherical waves pass the point x: the retarded one, created in the past τ_{ret}, and the advanced one, created in the future τ_{adv}. J is the source of both.

is a null four-vector, $K^2 = 0$, and represents a light-cone generator, a tangent to the light-cone. Constraint (8) is a condition of consistency of (6). It defines a family of hyperplanes tangent to the light-cone defined by $R^2 = 0$. Together, these two constraints require that x and $z(\tau_{\text{ret}})$ belong to a same straight line, generator of the light-cone with a vertex at the point x. This generator is the one tangent to K^μ, according to (9), and orthogonal to K_μ, according to (8): $K^\mu K_\mu = 0$. At the vertex of a light-cone the generators come in pairs: $K^\mu := (K^0, \boldsymbol{K})$ and $\bar{K}^\mu := (K^0, -\boldsymbol{K})$.

Together, (6) and (8) produce a much more restrictive causality constraint: a free massless physical object is restricted to remain on its light-cone generator (labelled by K). This is a very powerful restriction and radically changes the nature of field theory. One is no longer dealing with distributed fields defined on the whole light-cone but with a localized object (its (t = constant)-intersection is not a two-sphere but a point!) defined on a light-cone generator. Or, in other words, the part of a wavefront of $A(x)$ that moves along a light-cone generator must remain on this same generator. This is in direct contradiction to the idea behind the Huyghens' principle that each point of a wavefront acts as a secondary source emitting signal to all space directions; in other words, it assumes, at least in principle, that the signal at a point of a wavefront is made of contributions from all points of previous wavefronts. This idea could be appropriate for a description of light as a continuous wave manifestation, but not as a discrete one.

In contrast, constraints (6) and (8), together, imply from the start that a point on a wavefront propagates, on its light-cone generator, independently of all the other wavefront points. Each point of a wavefront, therefore, can be treated as an independent object by itself. In section 4 it will be shown that each point of a wavefront is created and annihilated (emitted and absorbed) not in a continuous way as is usually done in classical physics but in a discrete way, like a photon in quantum physics. It is thus justified to call each point of an electromagnetic wavefront a classical photon. One can associate the idea of a classical particle of null mass and dimensions to a classical photon. A classical photon is *related* to the intersection of (1), with a light-cone generator. The point is that (1) is a solution of (2) which describes a wave propagating on a light-cone. The appropriate description of a point propagating along a light-cone generator is done by an equation written in terms of the ∇ operator defined below. This is presented elsewhere [11], and will not be discussed here.

The simultaneous imposition of (6) and (8) then corresponds to an extended causality concept applied to massless objects; it is also readily extensible to massive objects [10]. It is appropriate for descriptions of particle-like fields with discrete interactions, that is, localized and propagating like a particle. Usually field theories are based on local causality, but it is possible to build a theory based on this extended causality [11].

Armed with these concepts of extend causality and classical photons one can present another physical interpretation of the above two LWSs. At the event x there are two classical photons. One, that was emitted by the electron current J, at $z(\tau_{\text{ret}})$ with $x^0 > z^0(\tau_{\text{ret}})$, and is moving in the K generator of the x-light-cone, $K^\mu := (K^0, \boldsymbol{K})$. J is its source. The other one, moving on a $\bar{K}$-generator, $\bar{K}^\mu := (K^0, -\boldsymbol{K})$, will be absorbed by J at $z(\tau_{\text{adv}})$, with $x^0 < z^0(\tau_{\text{adv}})$. J is its sink (see figure 2). They are both retarded solutions and correspond, respectively, to the creation and destruction of a classical photon. Exactly this: creation and destruction of particles in classical physics! This interpretation is only allowed with these concepts of extended causality and of classical photon; it is not possible with the continuous wave solutions. It will be instrumental for a clear understanding of how those one-century old problems of CED are worked out.

4. Energy tensor and integrability

When taking derivatives of $A(x)$, restriction (8), or equivalently, $K_\mu = -\frac{\partial\tau_{\text{ret}}}{\partial x^\mu}$ must be considered. This can turn a trivial calculation, for the untrained, into a mess. The best and more fruitful approach, in my opinion, is to take x and τ_{ret} as five independent parameters, and absorb restriction (8) in the definition of a new derivative operator ∇, replacing the usual one:

$$\frac{\partial}{\partial x^\mu} \Rightarrow \nabla_\mu := \frac{\partial}{\partial x^\mu} + \frac{\partial\tau}{\partial x^\mu}\frac{\partial}{\partial\tau} = \frac{\partial}{\partial x^\mu} - K_\mu\frac{\partial}{\partial\tau} \tag{10}$$

or $\nabla_\mu := \partial_\mu - K_\mu\partial_\tau$, by a shorter notation. Therefore, $\partial_\mu A(x)$, with the explicit restriction $|_{\tau_{\text{ret}}}$, is equivalent to $\nabla_\mu A(x)$ without any restrictions.

$$\partial_\mu A(x)|_{\tau_{\text{ret}}} = \nabla_\mu A(x, \tau). \tag{11}$$

This restriction, $|_{\tau_{\text{ret}}}$, is, therefore, implicitly assumed everywhere in this paper except when otherwise clearly stated, as for example in the following section when the points $\tau_{\text{ret}} \pm \mathrm{d}\tau$ are considered. The use of ∇, as defined in (11), simplifies the notation, as it is no longer necessary to carry this restriction $|_{\tau_{\text{ret}}}$.

The geometric meaning of ∇ is quite clear; it is the derivative allowed by restrictions (6) and (8), that is, displacements along the K light-cone generator only. One could complete the geometric picture seeing the operator ∇ as a kind of 'covariant derivative' with the

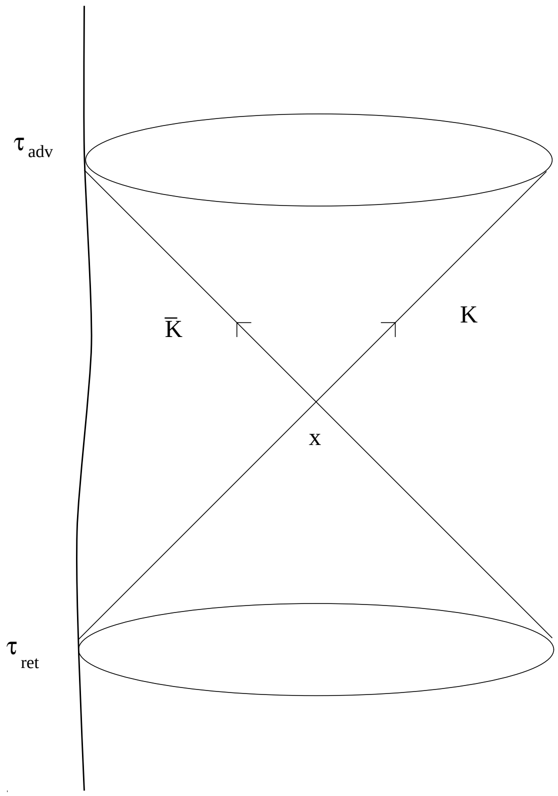

Figure 2. Creation and annihilation of particle in classical physics as a new interpretation of the LWS. At x there are two (classical) photons. One, created in the past by J, at $\tau_{\rm ret}$, and propagating along the light-cone generator K. J is its source. The other one, propagating along $\bar{K}$, will be absorbed in the future by J, at $\tau_{\rm adv}$. J is its sink. Both are retarded and point-like solutions.

connections of a new spacetime geometry [10] that would give a description equivalent to the old Minkowski spacetime plus generalizations of constraints (6) and (8). This would correspond to a complete geometrization of the extended causality concept.

It is funny that although the standard view of CED uses the concept of local causality (that is, only constraint (6)) for interpreting the LWSs, it actually does all further calculation (the Maxwell stress tensor, for example) according to the rules of the extended causality concept. In other words, the electromagnetic field obtained from (1) are the variations of (1) along a light-cone generator.

Therefore,

$$\nabla_\mu A^\nu = \nabla_\mu \frac{V^\nu}{\rho} = -\frac{K_\mu a^\nu}{\rho} - \frac{V^\nu}{\rho^2}\nabla_\mu \rho = -K_\mu \frac{a^\nu}{\rho} - \frac{V^\nu(K_\mu E - V_\mu)}{\rho^2} \tag{12}$$

with

$$E = 1 + aR = 1 + \rho a_K \tag{13}$$

as $\nabla_\mu V^\nu = -K_\mu a^\nu$ and

$$\nabla_\mu \rho = K_\mu E - V_\mu \tag{14}$$

where $a_K := aK$. Observe that the Lorentz gauge condition is automatically satisfied

$$\nabla A = -\frac{\rho a_K + V\nabla\rho}{\rho^2} = 0 \tag{15}$$

as $VK = -1$, $V^2 = -1$, and $V\nabla\rho = 1 - E = -\rho a_K$.

For simplicity we use a notation where $[A, B]$ stands for $[A_\mu, B_\nu] := A_\mu B_\nu - B_\mu A_\nu$ and (A, B) stands for $(A_\mu, B_\nu) := A_\mu B_\nu + A_\nu B_\mu$. The Maxwell field $F_{\mu\nu} := \nabla_\nu A_\mu - \nabla_\mu A_\nu$, is found to be

$$F = \frac{1}{\rho^2}[K, W] \tag{16}$$

with

$$W^\mu = \rho a^\mu + EV^\mu. \tag{17}$$

The electron self-field energy–momentum tensor, $4\pi\Theta = FF - \frac{\eta}{4}F^2$, is

$$4\pi\rho^4\Theta^{\mu\nu} = [K^\mu, W^\alpha][K_\alpha, W^\nu] - \frac{\eta^{\mu\nu}}{4}[K^\alpha, W^\beta][K_\beta, W_\alpha] \tag{18}$$

or in an expanded expression

$$4\pi\rho^4\Theta = (K, W) + KKW^2 + WWK^2 + \frac{\eta}{2}(1 - K^2W^2) \tag{19}$$

as $KW = -1$. The use of rather compact expressions such as (18) instead of (19) is preferred because besides being compact they will make the calculation of the zero-distance limit easier in the following sections. With $W^2 = \rho^2a^2 - E^2 = \rho^2a^2 - (1 + \rho a_K)^2$, Θ may be written, according to its powers of ρ, as $\Theta = \Theta_2 + \Theta_3 + \Theta_4$. If the K^2-terms are neglected then

$$4\pi\rho^2\Theta_2|_{K^2=0} = -KK(a^2 - a_K{}^2) \tag{20}$$

$$4\pi\rho^3\Theta_3|_{K^2=0} = 2KKa_K - (K, a + Va_K) \tag{21}$$

$$4\pi\rho^4\Theta_4|_{K^2=0} = KK - (K, V) - \frac{\eta}{2} \tag{22}$$

which are the usual expressions that one finds, for example in [1–3, 5, 6, 8]. Observe that

$$K\Theta_2|_{K^2=0} = 0 \tag{23}$$

which is important in the identification of Θ_2 with the radiated [5] part of Θ, and that

$$K\Theta_3|_{K^2=0} = 0. \tag{24}$$

The presence of non-integrable singularities in the electron self-field energy tensor is a major problem. $\Theta_2|_{K^2=0}$, although singular at $\rho = 0$, is nonetheless integrable. By that it is meant that it produces a finite flux through a spacelike hypersurface σ of normal n, that is, $\int \mathrm{d}^3\sigma\, \Theta_2 n$ exists [6], while $\Theta_3|_{K^2=0}$ and $\Theta_4|_{K^2=0}$ are not integrable; they generate, respectively, the problematic Schott term in the LDE and a divergent term, the electron bound four-momentum [5], which includes the so-called electron self-energy. Previous attempts, based on the distribution theory, for taming these singularities have relied on modifications of the Maxwell theory with addition of extra terms to $\Theta|_{K^2=0}$ on the electron worldline (see, for example [5, 6, 8]). They redefine $\Theta_3|_{K^2=0}$ and $\Theta_4|_{K^2=0}$ at the electron worldline in order to make them integrable without changing them at $\rho > 0$, so to preserve the standard results of CED. But this is always an *ad hoc* introduction of something strange to the theory. Another unsatisfactory aspect of this procedure is that it regularizes the above integral but leaves an unexplained and unphysical discontinuity in the flux of four-momentum, $\int \mathrm{d}x^4\, \Theta^{\mu\nu}\nabla_\nu\rho\delta(\rho - \varepsilon)$, through a cylindrical hypersurface $\rho = \varepsilon =$ constant enclosing the charge worldline. It is particularly interesting that, as it will be shown in the sequence, instead of adding anything one should actually not drop out the null K^2-terms. Their contribution (not null, in an appropriate limit) cancels the infinities. The same problem occurs in the derivations of the electron equation of motion from these incomplete expressions of Θ. The Schott term in the Lorentz–Dirac equation is a consequence; it does not appear in the equation when the full expression of Θ is correctly used. The point is that K and Θ are defined only for $\rho > 0$. $K^2 = 0$ is also true only for $\rho > 0$. Everybody in the literature does not use the complete expression (19) for Θ, but instead the shorter $\Theta|_{K^2=0}$-expressions when considering the limit of ρ tending to zero. Therefore, there is a generalized use of an implicit assumption that K^2 remains null at the limit $\rho = 0$. This is false, as is shown in section 5, and compromises all results in the actual literature.

5. The zero-distance limit

Θ is an explicit function of K and ρ. K is defined only for $\rho > 0$, $K := R/\rho$, and so is also $K^2 = 0$. At the limiting point $\rho = 0$ they produce an indeterminacy, as R necessarily also tends to zero: $(R \to 0)$ or $x \to z(\tau_{\text{ret}})$, along the light-cone generator K^μ. By force of the constraints (6) and (8), as x and $z(\tau_{\text{ret}})$ must remain on the same straight line, the light-cone generator K, the limit $\rho \to 0$ necessarily also implies on $x^\mu \to z(\tau_{\text{ret}})^\mu$ or $R^\mu \to 0$.

The indeterminacy of $K = \frac{R}{\rho}$ at $z(\tau_{\text{ret}})$, can be evaluated at neighbouring points $\tau = \tau_{\text{ret}} \pm \mathrm{d}\tau$ by the L'Hôpital's rule and $\frac{\partial}{\partial\tau}$ (see figure 3). This application of the L'Hôpital's rule then corresponds then to finding two simultaneous limits: $\rho \to 0$ and $\tau \to \tau_{\text{ret}}$.

As

$$\partial_\tau \rho \equiv \dot{\rho} = -(1 + aR) \tag{25}$$

and

$$\dot{R} = -V \tag{26}$$

then

$$\lim_{\substack{\rho\to 0\\ \tau\to\tau_{\text{ret}}}} K|_{\substack{R^2=0\\ R\,\mathrm{d}R=0}} = V. \tag{27}$$

This double limiting process is of course distinct of the single $(\rho \to 0)$-limit, which cannot avoid the singularity. For the simplicity of notation the use of just $\lim_{\rho\to 0}$ will be kept but always with the implicit meaning of this double limit as indicated in (27). For example, by

$$\lim_{\rho\to 0} K^2 = -1 \tag{28}$$

it is meant

$$\lim_{\substack{\rho\to 0\\ \tau\to\tau_{\text{ret}}}} K^2|_{\substack{R^2=0\\ R\,\mathrm{d}R=0}} = -1. \tag{29}$$

This invalidates all previous results in the literature on the CED short-distance behaviour because they have all been obtained from $\Theta|_{K^2}$ as it is valid only for $\rho > 0$; but then it could not be used in the $(\rho \to 0)$-limit. Besides, the correct limit (27) has not been used.

This limit (27) and the geometry behind it require a consistent physical interpretation that implies on a new connection between classical and quantum physics. The classical electromagnetic interaction between two point charges as described by the LWS $A(x)$, comprises the entire light-cone, $R^2 = 0$, that is, all the space surrounding each charge. But

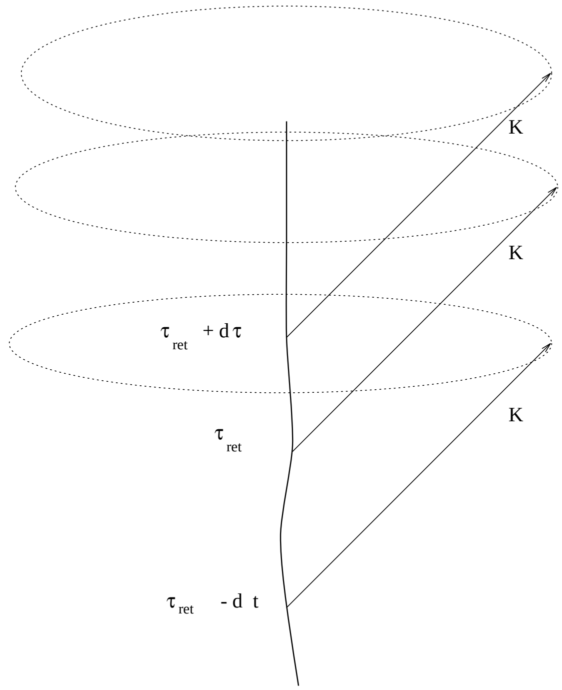

Figure 3. Double limiting process: $x \to z$ along K and $\tau \to \tau_{\rm ret}$. The single limiting process $x \to z(\tau_{\rm ret})$ along K (or $\rho \to 0$) does not solve the indeterminacy in the definition of K at $\rho = 0$. $K = \frac{x - z(\tau_{\rm ret})}{\rho}$ for $\rho > 0$. It takes a second simultaneous limit $\tau \to \tau_{\rm ret}$ along the electron worldline.

the simultaneous imposition of $R^2 = 0$ and $R\,\mathrm{d}R = 0$ (or $\mathrm{d}\tau + K\,\mathrm{d}x = 0$) implies that only the part of $A(x)$ contained in the light-cone generator, K, connecting the two charges must be considered at a time. This is the possible description, in classical physics, of the electromagnetic fundamental interaction: the exchange of a single photon. The light-cone generator is the photon classical trajectory (see figures 2 and 4). Now it is possible to understand the reasons of the $(\frac{0}{0})$-indeterminacy at $\tau = \tau_{\rm ret}$. In the limit of $\rho \to 0$ at $\tau = \tau_{\rm ret}$ there are three distinct velocities: K, the photon four-velocity, and V_1 and V_2, the electron initial and final four-velocities. The singularity at $\tau_{\rm ret}$ is not associated to any infinity but to an indeterminacy in the tangent of the electron worldline. At $\tau = \tau_{\rm ret} + \mathrm{d}\tau$ there is only V_2, and only V_1 at $\tau = \tau_{\rm ret} - \mathrm{d}\tau$. In other words, $\tau_{\rm ret}$ is an isolated singular point on the electron worldline; its neighbouring points $\tau_{\rm ret} \pm \mathrm{d}\tau$ are not singular. This is in flagrant contradiction to the CED assumption of a continuous emission process, because in this case, all points on the electron worldline would be singular points, like $\tau_{\rm ret}$. This completes the justification for the introduction of the classical photon concept: the part of the electromagnetic interaction contained in a light-cone generator is independent of the other parts contained in the other light-cone generators and, besides, it is discretely emitted and absorbed. There is a classical photon at $\tau_{\rm ret}$ but there is none at $\tau_{\rm ret} \pm \mathrm{d}\tau$. This picture will receive a further confirmation by the calculation of the energy flux from the charge at $z(\tau_{\rm ret} \pm \mathrm{d}\tau)$, in section 7. It is remarkable that one can find in a classical (Lienard–Wiechert) solution these traits of the quantum nature of the radiation emission process. They show a new bridge between classical and quantum field theories: the classical field is an effective representation of the effects of a photon exchange smeared in the charge light-cone. $R^2 = 0$ and $R\,\mathrm{d}R = 0$ establish an extended constraint of causality that retrieves the interaction one-photon-exchange character from the smeared-interaction field.

6. Some useful mathematical tools

To find this double limit of something when $\rho \to 0$ and $\tau \to \tau_{\mathrm{ret}}$ will be done so many times in this paper that it is better to do it in a more systematic way. One wants to find

$$\lim_{\rho\to 0} \frac{N(R,\ldots)}{\rho^n} \tag{30}$$

where $N(R,\ldots)$ is a homogeneous function of R, $N(R,\ldots)|_{R=0} = 0$. Then, one has to apply the L'Hôpital's rule consecutively until the indeterminacy is resolved. As $\frac{\partial\rho}{\partial\tau} = -(1+aR)$, the denominator of (30) at $R=0$ will only be different to zero after the nth-application of the L'Hôpital's rule, and then, its value will be $(-1)^n n!$.

If p is the smallest integer such that $N(R,\ldots)_p|_{R=0} \neq 0$, where $N(R)_p := \frac{\mathrm{d}^p}{\mathrm{d}\tau^p} N(R,\ldots)$, then

$$\lim_{\rho\to 0} \frac{N(R,\ldots)}{\rho^n} = \begin{cases} \infty & \text{if } p < n \\ (-1)^n \dfrac{N(0,\ldots)_p}{n!} & \text{if } p = n \\ 0 & \text{if } p > n. \end{cases} \tag{31}$$

- Example 1: $\begin{cases} K = \frac{R}{\rho} & n = p = 1 \Longrightarrow \lim_{\rho\to 0} K = V \\ K^2 = \frac{R\eta R}{\rho^2} & n = p = 2 \Longrightarrow \lim_{\rho\to 0} K^2 = -1. \end{cases}$
- Example 2: $\frac{[K,a]}{\rho} = \frac{[R,a]}{\rho^2} \Longrightarrow p = 1 < n = 2 \Longrightarrow \lim_{\rho\to 0} \frac{[K,a]}{\rho}$ diverges.
- Example 3: $\frac{a_K}{\rho}[K,V] = -\frac{aR}{\rho^3}[R,V] \Longrightarrow p = 4 > n = 3$, $\lim_{\rho\to 0} \frac{a_K}{\rho}[K,V] = 0$.
- Example 4: $\frac{[K,V]}{\rho^2} = \frac{[R,V]}{\rho^2} \Longrightarrow p = 2 < n = 3 \Longrightarrow \lim_{\rho\to 0} \frac{[K,V]}{\rho^2}$ diverges.

Finding these limits for more complex functions can be made easier with two helpful expressions,

$$N_p = \sum_{a=0}^{p} \binom{p}{a} A_{p-a} B_a \tag{32}$$

and

$$N_p = \sum_{a=0}^{p} \sum_{c=0}^{a} \binom{p}{a} \binom{a}{c} A_{p-a} B_{a-c} C_c \tag{33}$$

valid when $N(R)$ has, respectively, the forms $N_0 = A_0 B_0$, or $N_0 = A_0 B_0 C_0$, where A, B and C represent possibly distinct functions of R, and the subindices indicate the order of $\mathrm{d}/\mathrm{d}\tau$. For example: $A_0 = A$; $A_1 = \partial_\tau A$; $A_2 = \partial_\tau^2 A$, etc. So, for using (31)–(33), one just has to find the τ-derivatives of A, B and C that produce the first non-zero term at the point limit of $R \to 0$.

Consecutive derivatives of products of complex functions can become unwieldy. So it is worth introducing the concept of 'τ-order' of a function, meaning the lowest order of the τ-derivative of a function that produces a non-zero result at the limiting point $R = 0$. Let $\mathcal{O}[f(x)]$ represent the 'τ-order' of f(x). So, for example, from (25) and (26) one sees that

$$\mathcal{O}[R] = 1 \tag{34}$$

$$\mathcal{O}[\rho] = 1. \tag{35}$$

As $\partial_\tau(aR) = -\dot{a}R$ and $\partial_\tau^2(aR) = -\ddot{a}R - \dot{a}V = -\ddot{a}R + a^2 = a^2 + \mathcal{O}(R)$, then

$$\mathcal{O}[aR] = 2. \tag{36}$$

For finding the N_p of (32) and (33) it is then necessary to consider only the terms with the lowest τ-order on each factor.

Some combinations of terms have derivatives that cancel parts of each other resulting in a higher τ-order term. For example,

$$\begin{aligned}
&\partial_\tau(R^2 + \rho^2) = +2\rho - 2\rho E = -2\rho aR\\
&\partial_\tau^2(R^2 + \rho^2) = 2EaR - 2\rho\dot{a}R = 2(aR - \rho\dot{a}R) + \mathcal{O}(R^4)\\
&\partial_\tau^3(R^2 + \rho^2) = 2(\dot{a}R + E\dot{a}R - \rho a^2) + \mathcal{O}(R^3) = 4\dot{a}R - 2\rho a^2 + \mathcal{O}(R^3)\\
&\partial_\tau^4(R^2 + \rho^2) = 4a^2 + 2a^2 + \mathcal{O}(R^2) = 6a^2 + \mathcal{O}(R^2).
\end{aligned}$$

So,

$$\mathcal{O}[R^2 + \rho^2] = 4$$

although

$$\mathcal{O}[R^2] = \mathcal{O}[\rho^2] = 2.$$

Observe that one has to care only with the lowest τ-order terms as the other ones, grouped in $\mathcal{O}(R)$, will not survive the limit $R \to 0$. Also, it is not necessary to write the τ-derivatives of factors that will not reduce its τ-order. For example in

$$\partial_\tau(RV + \mathcal{O}(R^2)) = -VV + \mathcal{O}(R^2)$$

the term Ra was absorbed in $\mathcal{O}(R^2)$. In this way we avoid writing and taking unnecessary derivatives of long expressions with terms that will not contribute to the final result.

An important property of $\mathcal{O}[f(x)]$:

$$\mathcal{O}[ABC] = \mathcal{O}[A] + \mathcal{O}[B] + \mathcal{O}[C] \tag{37}$$

so that, p as defined by (31) and (33), is

$$p = \mathcal{O}[ABC] = \mathcal{O}[A] + \mathcal{O}[B] + \mathcal{O}[C]. \tag{38}$$

7. Fluxes and equation of motion

The motion of a classical electron [1–3] is described by the Lorentz–Dirac equation,

$$ma = F_{\text{ext}}V + \tfrac{2}{3}(\dot{a} - a^2V) \tag{39}$$

where m is the electron mass and F_{ext} is an external electromagnetic field. The presence of the Schott term, $\frac{2}{3}\dot{a}$, is the cause of all of its pathological features, such as microscopic non-causality, runaway solutions, pre-acceleration, and other bizarre effects [4]. On the other hand, its presence is apparently necessary for the energy–momentum conservation; without it, it would require a contradictory null radiance for an accelerated charge, as $aV = 0$, $VF_{\text{ext}}V = 0$ and $\dot{a}V + a^2 = 0$. This makes the Lorentz–Dirac equation the greatest paradox of classical field theory as it cannot simultaneously preserve both the causality and the energy conservation [1–3].

The Lorentz–Dirac equation can be obtained from energy–momentum conservation, which leads to

$$\int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, (F^{\mu\nu}_{\mathrm{ext}} V_\nu - ma^\mu) = \lim_{\substack{\varepsilon_1\to 0\\ \varepsilon_2\to\infty}} \int_{\mathcal{V}} \mathrm{d}x^4\, \nabla_\nu \Theta^{\mu\nu} \theta(\varepsilon_2, \rho, \varepsilon_1)\theta(\tau_2, \tau, \tau_1)$$
$$= \lim_{\substack{\varepsilon_1\to 0\\ \varepsilon_2\to\infty}} \int_{\mathcal{V}} \mathrm{d}x^4\, \Theta^{\mu\nu} \{\nabla_\nu \rho(\theta(\rho-\varepsilon_1)\delta(\varepsilon_2-\rho)$$
$$-\delta(\rho-\varepsilon_1)\theta(\varepsilon_2-\rho))\theta(\tau_2,\tau,\tau_1)$$
$$+\nabla_\nu \tau(\delta(\tau_2-\tau)\theta(\tau-\tau_1) - \theta(\tau_2-\tau)\delta(\tau-\tau_1))\theta(\varepsilon_2,\rho,\varepsilon_1)\} \qquad (40)$$

where $\theta(a_2, x, a_1) = \theta(a_2 - x)\theta(x - a_1)$, the product of two Heaviside functions, and τ_2, τ_1, ε_2 and ε_1 are constants, with $\tau_2 > \tau_1$ and $\varepsilon_2 > \varepsilon_1$. $\theta(\tau_2, \tau, \tau_1) = \theta(\tau_2 - \tau)\theta(\tau - \tau_1)$ defines the spacetime region between the two light-cones of vertices at τ_2 and τ_1. The product of these four Heaviside functions defines the closed boundary of an hypervolume that is totally inside the integration domain $\mathcal{V}$. The passage from the first to the second and third lines of (40) involves integration by parts, the divergence theorem and the use of

$$\Theta^{\mu\nu}\theta(\varepsilon_2, \rho, \varepsilon_1)\theta(\tau_2, \tau, \tau_1)|_{\partial\mathcal{V}} = 0. \qquad (41)$$

$\nabla_\nu \Theta^{\mu\nu} = 0$ in the hypervolume, for $\varepsilon_1 > 0$, assures that the integral on the RHS of the first line of (40) is null for any $\varepsilon_1 > 0$, but not, as it will be shown now, in the limit when ε_1 tends to zero. This approach is equivalent to one where $\Theta^{\mu\nu}$ is treated as a distribution [6, 8]. Both are equally rigourous and give the same results, but this one is simpler as it dispenses the use of a compact test function, which is replaced by $\theta(\varepsilon_2 - \rho)$, in its role of allowing a compact domain of integration. No infinity appears in this approach and so it is not necessary to consider the distribution character of $\Theta^{\mu\nu}$. The terms in the second and third lines of (40) are the fluxes of energy–momentum through the respective hypersurfaces $\rho = \varepsilon_1$, $\rho = \varepsilon_2$, $\tau = \tau_2$ and $\tau = \tau_1$. They are well known in the literature [1, 7], for $\varepsilon_1 > 0$. The flux on a Bhabha tube $\rho = \varepsilon > 0$ is given by

$$\Phi_1(\varepsilon)^\mu = \int \mathrm{d}x^4\, \Theta^{\mu\nu} \nabla_\nu \rho \delta(\varepsilon - \rho)\theta(\tau_2, \tau, \tau_1). \qquad (42)$$

With

$$W\nabla\rho \equiv 0 \qquad (43)$$
$$KW = -1 \qquad (44)$$
$$K\nabla\rho = 1 + K^2 E \qquad (45)$$

and $K^2 = 0$, as $\varepsilon > 0$, in (18) one has

$$4\pi\rho^4 \Theta\nabla\rho = W + K\left(W^2 + \frac{E}{2}\right) - \tfrac{1}{2}V = \rho a + V(\rho a_K + \tfrac{1}{2})$$
$$+K(\rho^2(a^2 - a_K^2) - \tfrac{1}{2}(1 + 3\rho a_K)). \qquad (46)$$

Now using retarded coordinates [5, 6, 12] where $\mathrm{d}^4x = \mathrm{d}\tau\, \rho^2\, \mathrm{d}\rho\, \mathrm{d}^2\Omega$, and

$$\frac{1}{4\pi}\int \mathrm{d}^2\Omega\, K^\alpha = V^\alpha \qquad (47)$$
$$\frac{1}{4\pi}\int \mathrm{d}^2\Omega\, K^\alpha K^\beta = \frac{1}{3}\Delta^{\alpha\beta} + V^\alpha V^\beta \qquad (48)$$
$$\frac{1}{4\pi}\int \mathrm{d}^2\Omega\, K^\alpha K^\beta K^\gamma = \Delta^{(\alpha\beta} V^{\gamma)} + V^\alpha V^\beta V^\gamma \qquad (49)$$

where $\Delta = \eta + VV$, and the parenthesis on the superscripts mean total symmetrization, one has for (42)

$$\Phi_1(\varepsilon)^\mu = \int_{\tau_1}^{\tau_2} \mathrm{d}\tau \left(\frac{2}{3}a^2 V^\mu - \frac{a^\mu}{2\varepsilon}\right). \tag{50}$$

The total flux on the sections $0 < \varepsilon_1 < \rho < \varepsilon_2$ of the light-cones $\tau = \tau_2$ and $\tau = \tau_1$ is given by

$$\begin{aligned}\Phi_2(\varepsilon_2)^\mu - \Phi_2(\varepsilon_1)^\mu &= -\int \mathrm{d}x^4\, \Theta^{\mu\nu} K_\nu \theta(\varepsilon_2, \rho, \varepsilon_1)(\delta(\tau_2 - \tau) - \delta(\tau - \tau_1))\\ &= \frac{1}{2}(V(\tau_2)^\mu - V(\tau_1)^\mu)\left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1}\right) = \frac{1}{2}\int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, a^\mu \left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1}\right)\end{aligned} \tag{51}$$

as $\nabla_\nu \tau = -K_\nu$ and $4\pi\rho^4\Theta K = \frac{1}{2}K$, for $\rho > 0$. In the notation used in (51), $\Phi_2(\varepsilon_2)$ and $\Phi_2(\varepsilon_1)$ are, respectively, the upper and the lower limit of the ρ-integration in (51):

$$\Phi_2(\varepsilon_2) = \frac{1}{2}\int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, \frac{a}{\varepsilon_2} \tag{52}$$

and

$$\Phi_2(\varepsilon_1) = \frac{1}{2}\int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, \frac{a}{\varepsilon_1}. \tag{53}$$

One then sees from (42) and (51) that the integral on the RHS of the first line of (40) is equal to

$$\int \mathrm{d}x^4\, \nabla_\nu \Theta^{\mu\nu} \theta(\varepsilon_2, \rho, \varepsilon_1)\theta(\tau_2, \tau, \tau_1) = \Phi_1(\varepsilon_2) - \Phi_1(\varepsilon_1) + \Phi_2(\varepsilon_2) - \Phi_2(\varepsilon_1) \tag{54}$$

and that this (the RHS) is identically null, for any $\varepsilon_1 > 0$, which is in agreement with $\nabla_\nu \Theta^{\mu\nu}$ being null at $\rho > 0$. These results can be extended to $\varepsilon_2 \to \infty$ but not to $\varepsilon_1 \to 0$ because of the explicit dependence of $\Theta^{\mu\nu}$ on the null four-vector K, which is defined only for $\rho > 0$. At $\rho = 0$ its definition ($K = \frac{R}{\rho}$) gives an indeterminacy. In the literature it is implicitly assumed that K remains a null four-vector at the limiting $\rho = 0$. Besides not being correct, as it has been shown, this produces a diverging flux (the self-energy problem) and the controversial Schott term in the Lorentz–Dirac equation.

In the $(\varepsilon_2 \to \infty)$-limit one has $\Phi_2(\varepsilon_2 = \infty) = 0$ and $\Phi_1(\varepsilon_2 = \infty) = \int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, \frac{2}{3}a^2 V^\mu$. Therefore, with (42) and (51), equation (40) can be written as

$$\int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, (F_{\text{ext}}^{\mu\nu} V_\nu - m a^\mu) = -\lim_{\varepsilon_1 \to 0}\{\Phi_2(\varepsilon_1)^\mu + \Phi_1(\varepsilon_1)^\mu\} + \int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, \tfrac{2}{3}a^2 V^\mu. \tag{55}$$

Equations (31)–(33) will now be used to find the $(\varepsilon_1 \to 0)$-limit of $\Phi_1(\varepsilon_1) = \int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, \rho^2\, \mathrm{d}\rho\, \mathrm{d}^2\Omega\, \Theta\nabla\rho\delta(\rho - \varepsilon_1)$. But in (18), the definition of Θ, the second term is the trace of the first one and so one just has to consider this last one because the behaviour of its trace under this limiting process can then easily be inferred. So, as $K = R/\rho$, and $\nabla\rho = (KE - V)$, one has schematically for the first term of (18) in $\rho^2\Theta\nabla\rho$,

$$\lim_{\rho\to 0} \frac{N(R, \ldots)}{\rho^n} = \lim_{\rho\to 0} \frac{\rho^2[K, W][K, W](KE - V)}{\rho^4} = \lim_{\rho\to 0} \frac{[R, W][R, W](RE - V\rho)}{\rho^5}. \tag{56}$$

Then, from the comparison with (30) and (33),

$$
\begin{aligned}
&A_0 = B_0 = [R, W] = [R, a\rho + VE] = [R, a\rho + V] + \mathcal{O}(R^3)\\
&A_1 = B_1 = [-V, \rho a + V] + [R, -aE + a] + \mathcal{O}(R^2) = -[V, \rho a] + \mathcal{O}(R^2) \qquad (57)\\
&A_2 = B_2 = -[a, V] + \mathcal{O}(R)
\end{aligned}
$$

$$
\begin{aligned}
&C_0 = RE - V\rho = R - V\rho + \mathcal{O}(R^3)\\
&C_1 = -V - a\rho + VE + \mathcal{O}(R^2) = -a\rho + \mathcal{O}(R^2) \qquad (58)\\
&C_2 = a + \mathcal{O}(R).
\end{aligned}
$$

Therefore, for producing a possibly non-zero N_p, according to (33), a, c and p must be given by

$$
\begin{aligned}
&c = 2\\
&p - a = a - c = 2 \Longrightarrow p = 6 > n = 5.
\end{aligned}
$$

Or in a briefer way,

$$
\begin{aligned}
&\mathcal{O}[[R, W]] = 2\\
&\mathcal{O}[RE - V\rho] = 2
\end{aligned}
$$

and then, using (38),

$$p = 2\mathcal{O}[[R, W]] + \mathcal{O}[RE - V\rho] = 6 > n = 5.$$

Therefore,

$$\lim_{\varepsilon_1 \to 0} \Phi_1(\varepsilon_1) = 0. \qquad (59)$$

The flux of energy and momentum of the electron self-field through the $(\rho = \varepsilon_1)$-hypersurface in (40) is null at $\varepsilon_1 = 0$. This is a new result, a consequence of (27). In the standard approach, with the uncomplete expressions of $\Theta^{\mu\nu}$, the contribution from this term produces the problematic Schott term and a diverging expression, the electron bound-momentum which requires mass renormalization [9]. In (59) if one had used the K^2-terms expurgated energy tensor, which is the one used in the literature, one would have found an infinity on its RHS, even using (27). The K^2-terms in (18) cancel the infinities.

For the evaluation of $\lim_{\varepsilon_1 \to 0} \Phi_2(\varepsilon_1)$ one finds from (19) and $KW = -1$ that

$$4\pi\rho^2\Theta^{\mu\nu}K_\nu = \frac{K^\mu}{2\rho^2}(1 - K^2W^2) = \frac{K^\mu}{2}\left\{\frac{1 + K^2}{\rho^2} - 2\frac{K^2 aK}{\rho} - K^2(a^2 - a_K{}^2)\right\} \qquad (60)$$

and

$$4\pi\int \mathrm{d}\rho\, \rho^2\Theta^{\mu\nu}K_\nu = -\frac{K^\mu}{2}\left\{\frac{1 + K^2}{\rho} + 2K^2 aK \ln\rho + K^2(a^2 + a_K{}^2)\rho\right\}. \qquad (61)$$

But, again from comparison with (30)–(33),

$$\lim_{\rho \to 0} \frac{K^\mu(1 + K^2)}{\rho} = \frac{R^\mu(\rho^2 + R^2)}{\rho^4} = 0 \qquad (62)$$

as $\mathcal{O}[R] + \mathcal{O}[\rho^2 + R^2] = 1 + 4 > 4$.

Also, as

$$\mathcal{O}[K^\mu K^2 aK] = 1 \qquad (63)$$

one can say from (35) that

$$\lim_{\rho \to 0} K^\mu K^2 aK \sim \lim_{\rho \to 0} \rho \qquad (64)$$

that is, $K^\mu K^2 aK$ tends to zero with ρ as fast as ρ. So,

$$\lim_{\rho\to 0} K^\mu K^2 aK \ln\rho = \lim_{\rho\to 0} \rho \ln\rho = 0. \tag{65}$$

The last term of (61) is also null in the $(\rho \to 0)$-limit, and so one can conclude that

$$\lim_{\rho\to 0} \Phi_2(\varepsilon_1) = 0. \tag{66}$$

The meaning of (59) and (66): at $\rho = 0$ and $\tau = \tau_{\rm ret} \pm {\rm d}\tau$ there is only the electron! No self-field, no photon! The flux from the charge is zero for $\tau_{\rm ret} \pm {\rm d}\tau$ and, of course, non-zero at $z(\tau_{\rm ret})$. This confirms the picture of a discrete radiation process. It takes the limiting $\rho \to 0$ to be seen because at $\rho > 0$ it is masqueraded by the field average character, as will be discussed later. This is in contradiction to the Gauss's law! It requires a revision of its physical meaning and of the Maxwell–Faraday concept of field, which will be done in section 9. First, one should discuss the issue of the electron equation of motion which will make the inadequacy of the picture of a continuous interaction in a short-distance scale more evident.

8. An effective equation of motion

With (59) and (66) in (55) one could write the electron equation of motion as

$$ma^\mu - F^{\mu\nu}_{\rm ext} V_\nu = -\tfrac{2}{3} a^2 V^\mu \tag{67}$$

but it is well known that this could not be a correct equation because it is not self-consistent: its LHS is orthogonal to V,

$$maV = 0 \qquad \text{and} \qquad VF_{\rm ext}V = 0 \tag{68}$$

while its RHS is not,

$$-\tfrac{2}{3} a^2 VV = \tfrac{2}{3} a^2. \tag{69}$$

This seems to be paradoxical until one has a clearer idea of what is happening. One must return to equation (40), where there is a subtle and very important distinction between its LHS and RHS. Its LHS is entirely determined by the electron instantaneous position, $z(\tau)$, while its RHS is determined by the sum of contributions from the electron self-field at all points. The equation of motion is the mathematical description of momentum conservation in the interaction. The LHS of (40) describes, therefore, the change of the electron momentum at a point (the electron instantaneous position) while the RHS describes the momentum carried away by the electron self-field which is distributed over the whole space. This is a consequence of the imposed dichotomic treatment: while the electron is described as a discrete and well localized object, a particle, its self-field is a non-localized object distributed over the entire space and whose contribution to the changes in the electron must be computed from all these points. It introduces a strong non-locality and excludes the possibility of a true equation of motion which would give an essentially local description. A true (in the sense of local) equation of motion for a classical charged particle is then possible only in the context of discrete interactions mediated by exchanged (classical) photons [11]. The RHS of (67) would be replaced in this case by the momentum of the emitted photon, while the LHS remains local as it is always defined at a single point, the electron position. The space-time average of this (then local) equation would reproduce (67). $F^{\mu\nu}_{\rm ext} V_\nu$ on the LHS of (67) is the spacetime average of the momentum exchanged between the electron and external charges while a^μ is the electron average acceleration. In the context of the

LWS (1), equation (67) must, therefore, be regarded as an effective equation that would be better represented as

$$ma^{\mu} - F^{\mu\nu}_{\text{ext}} V_{\nu} = -\langle \tfrac{2}{3} a^2 V^{\mu} \rangle \tag{70}$$

where the term in brackets represents the contribution from the electron self-field:

$$\langle \tfrac{2}{3} a^2 V^{\mu} \rangle = \lim_{\substack{\varepsilon_1 \to 0 \\ \varepsilon_2 \to \infty}} \int \mathrm{d}x^3 \, \nabla_{\nu} \Theta^{\mu\nu} \theta(\rho - \varepsilon_1) \theta(\varepsilon_2 - \rho). \tag{71}$$

This is more than just a change of notation; it explicitly implies on a clear distinction between the V inside and the V outside the bracket in (70):

$$\langle V \rangle \neq V.$$

This distinction between the LHS and RHS of (40) is missing in equation (67); it was deleted by the integration process. It represents the strong non-locality introduced at the beginning with the hypothesis of a continuous interacting field (1).

It makes no sense, therefore, to multiply (67) or (70) by V. This would be a mixing of instantaneous and average values. One should instead try to follow the associated physical picture. The LHS of (40) multiplied by V is null because the force that drives the electron with the four-velocity V delivers a power $ma_0 V^0$ that is equal to the work per unit time ($m\boldsymbol{a}\boldsymbol{V}$) realized by this force along the $\boldsymbol{V}$ direction (this, as is well known, is the physical meaning of $maV = 0$). But this reasoning does not apply to the RHS of (40) multiplied by V because the flux of radiated energy is through a spherical surface $\rho = \varepsilon_2$, along K at each point, not along V (except at $\rho = 0$, because of (27)); in order to make sense, as one is doing a balance of the flux rate of energy, one has to add this flux rate from each element of the integration domain. Based on considerations of symmetry one can anticipate that the final result must be null: to each point of a spherical hypersurface $\rho =$ constant, $\tau = \tau_2$, that gives a non-zero contribution there is another point giving an equal but with opposite sign contribution. The RHS of (67) cannot be used for this point-to-point calculation as it just represents a kind of average or resulting value. For this balance one must start again from the beginning. Contributions from the electron self-field must always be calculated through this point-by-point summation, like on the RHS of (40) for the flux of electromagnetic energy–momentum, through the walls of a Bhabha tube around the charge worldline, in the limit of $\rho \to 0$. In particular,

$$\int_{\tau_1}^{\tau_2} \mathrm{d}\tau \, (ma - V F_{\text{ext}}) V = - \lim_{\substack{\varepsilon_1 \to 0 \\ \varepsilon_2 \to \infty}} \int \mathrm{d}^4 x \, X_{\mu} \nabla_{\nu} \Theta^{\mu\nu} \theta(\varepsilon_2, \rho, \varepsilon_1) \theta(\tau_2, \tau, \tau_1) \tag{72}$$

where

$$X = \begin{cases} K & \text{if } \rho > 0 \\ V & \text{if } \rho = 0. \end{cases} \tag{73}$$

X, on the RHS of (72), gives the direction of the flux rate of the radiated energy; on the LHS this direction is given by V. Observe that $X(\tau_{\text{ret}})$ is x-dependent and so it does not commute with $\int \mathrm{d}^4 x$, that is, X inside and X outside the integral on the RHS of (72) give distinct results and, based on the above arguments, one is saying that (72) shows the correct way. Its LHS is, of course, null. It will now be shown that the RHS is also null, so that there is no longer any contradiction. One knows that

$$\nabla_{\nu} \Theta^{\mu\nu} = \frac{1}{4\pi} F^{\mu}{}_{\alpha} \nabla_{\nu} F^{\alpha\nu} = \frac{1}{4\pi} F^{\mu}{}_{\alpha} \, \Box \, A^{\alpha} \tag{74}$$

and by direct calculation one finds that

$$\Box A^{\mu} = \frac{K^2}{\rho^3}(3\rho E a^{\mu} + \rho^2 \dot{a}^{\mu} + (3E^2 + \rho^2 \dot{a}_K)V^{\mu}). \tag{75}$$

So, the integrand on the RHS of (72) is null for $\rho > 0$ as $K^2 = 0$ there. For simplicity one could then just have used V instead of X in (72), but see the next section for an alternative illuminating calculation. Therefore, one just has to verify that $\rho^2 V_{\mu}\nabla_{\nu}\Theta^{\mu\nu}|_{\rho=0}$ is finite, or equivalently that $\rho^3 V_{\mu}\nabla_{\nu}\Theta^{\mu\nu}|_{\rho=0} = 0$. As

$$V_{\mu}F^{\mu\nu} = \frac{1}{\rho^2}(EK^{\alpha} - W^{\alpha}) \tag{76}$$

then

$$4\pi\rho^5 V_{\mu}\nabla_{\nu}\Theta^{\mu\nu} = -K^2(2E\rho^2 a^2 + 3E(1-E^2) + \rho^2(\rho\dot{a}a - E\dot{a}_K)) \tag{77}$$

and

$$\lim_{\rho\to 0}\rho^3 V_{\mu}\nabla_{\nu}\Theta^{\mu\nu} = \lim_{\rho\to 0}\frac{R^2(2E\rho^2 a^2 + 3E(1-E^2) + \rho^2(\rho\dot{a}a - E\dot{a}_K))}{\rho^4} \tag{78}$$

and this is null at the limit $\rho \to 0$ because

$$\mathcal{O}[2E\rho^2 a^2 + 3E(1-E^2) + \rho^3\dot{a}a - \rho E\dot{a}_R] + \mathcal{O}[R^2] = 3 + 2 > 4$$

according to (32). So, both sides of (72) are equally null and there is no contradiction. This is in agreement with the fact that due to (2), (4) and to the antisymmetry of F,

$$V_{\mu}\nabla_{\nu}\Theta^{\mu\nu} = \frac{1}{4\pi}V_{\mu}F^{\mu}{}_{\alpha}\nabla_{\nu}F^{\alpha\nu} = V_{\mu}F^{\mu}{}_{\alpha}J^{\alpha} = 0.$$

9. Using the divergence theorem

For the sake of a better understanding of the meaning of X in equation (72) its RHS will be worked out by using the divergence theorem. Then one has

$$\begin{aligned}\int_{\tau_1}^{\tau_2}\mathrm{d}\tau\,(ma - VF_{\mathrm{ext}})V = \lim_{\substack{\varepsilon_1\to 0\\ \varepsilon_2\to\infty}}\int \mathrm{d}x^4\,\{&\Theta^{\mu\nu}\nabla_{\nu}X_{\mu}\theta(\varepsilon_2,\rho,\varepsilon_1)\theta(\tau_2,\tau,\tau_1)\\ &+X_{\mu}\Theta^{\mu\nu}[\nabla_{\nu}\rho(\delta(\rho-\varepsilon_1)\theta(\varepsilon_2-\rho)-\theta(\rho-\varepsilon_1)\delta(\varepsilon_2-\rho))\theta(\tau_2,\tau,\tau_1)\\ &-K_{\nu}(\delta(\tau-\tau_1)\theta(\tau_2-\tau)-\theta(\tau-\tau_1)\delta(\tau_2-\tau))\theta(\varepsilon_2,\rho,\varepsilon_1)]\}.\end{aligned} \tag{79}$$

The explicit dependence on $\nabla_{\nu}X_{\mu}$ makes it clear why one cannot just use K instead of X in (72): although $\lim_{\rho\to 0}K = V$, $\lim_{\rho\to 0}\nabla K \neq \nabla V = -Ka$.

For working out the first term on the RHS of (79) one needs (43)–(45) and

$$\nabla_{\mu}K_{\nu} = \nabla_{\mu}\left(\frac{R_{\nu}}{\rho}\right) = \frac{\eta_{\mu\nu} + K_{\mu}V_{\nu}}{\rho} - \frac{K_{\nu}}{\rho}\nabla_{\mu}\rho \tag{80}$$

$$\Theta^{\mu\nu}\eta_{\mu\nu} = 0. \tag{81}$$

Then, from (19) and $K^2 = 0$ one has for the upper limit

$$\lim_{\varepsilon_2\to\infty}\int^{\varepsilon_2}\mathrm{d}x^4\,\Theta^{\mu\nu}\nabla_{\nu}K_{\mu} = \lim_{\varepsilon_2\to\infty}\int\mathrm{d}\tau\int^{\varepsilon_2}\frac{\mathrm{d}\rho}{\rho^3} = 0. \tag{82}$$

For the lower limit $\nabla_{\nu}X_{\mu} = \nabla_{\nu}V_{\mu} = -K_{\nu}a_{\mu}$ and then, from (19),

$$4\pi\rho^4 K\Theta a = a_K(K^2W^2 - 1) = a_K(K^2+1) + \rho^2 a_K K^2(a^2 - a_K^2) - \rho a_K^2 K^2. \tag{83}$$

So,

$$\lim_{\rho\to 0} 4\pi \int \rho^2 \mathrm{d}\rho\, K\Theta a = -\frac{a_K(K^2+1)}{\rho} + a_K K^2(a^2 - a_K^2)\rho - a_K^2 K^2 \ln\rho = 0 \tag{84}$$

because

$$\lim_{\rho\to 0} \frac{a_K(K^2+1)}{\rho} = \lim_{\rho\to 0} \frac{a_R(R^2+\rho^2)}{\rho^4} = 0 \tag{85}$$

as

$$\mathcal{O}[a_R] + \mathcal{O}[R^2+\rho^2] = 2 + 4 > 4 \tag{86}$$

and

$$\lim_{\rho\to 0} a_K K^2(a^2 - a_K^2)\rho = 0. \tag{87}$$

For evaluating the limit of the last term of (84) one has considered that

$$K^2 a_K^2 = \frac{R^2(aR)^2}{\rho^4} \qquad \mathcal{O}[K^2 a_K^2] = 2 = \mathcal{O}[\rho^2] \tag{88}$$

to see that

$$\lim_{\rho\to 0} a_K^2 K^2 \ln\rho \sim \lim_{\rho\to 0} \rho^2 \ln\rho = 0. \tag{89}$$

It is important to use the appropriate values of X to have consistent results. The use, for example, of $X = V$ in the upper limit or of $X = K$ in the lower limit would produce inconsistent results. The second line of (79) is composed of two terms, with $\rho = \varepsilon_1$ and $\rho = \varepsilon_2$, respectively. For the $\rho = \varepsilon_1$ term, $X = V$ in the limit and then one has from (19) that

$$4\pi\rho^4 V\Theta\nabla\rho = (1 + K^2 E)(W^2 + E) + \frac{\rho a_K}{2}(1 - K^2 W^2). \tag{90}$$

Therefore,

$$\lim_{\rho\to 0} 4\pi\rho^2 V\Theta\nabla\rho = \lim_{\rho\to 0}\left(\frac{(\rho^2 + R^2 + R^2 a_R)(\rho^2 a^2 - a_R^2 - a_R)}{\rho^4} + \frac{a_R[\rho^2 + R^2 - R^2(\rho^2 a^2 - a_R^2 - 2a_R)]}{2\rho^4}\right) = 0 \tag{91}$$

because

$$\mathcal{O}[\rho^2 + R^2 + R^2 a_R^2] + \mathcal{O}[\rho^2 a^2 - a_R^2 - a_R] = 4 + 2 > 4 \tag{92}$$

and

$$\mathcal{O}[a_R] + \mathcal{O}[(\rho^2 + R^2) - R^2(\rho^2 a^2 - a_R^2 - 2a_R)] = 2 + 4 > 3. \tag{93}$$

Again one only has consistent results if one uses the correct values of X in its respective limiting situation.

For the $\rho = \varepsilon_2$ term one has $X = K$ and then

$$4\pi\rho^2 K\Theta\nabla\rho = \frac{1}{2}\frac{K\nabla\rho}{\rho^2} = -\frac{1}{2\rho^2}.$$

So, it is null in the limit when $\rho \to \infty$.

Finally, the third line of (79) does not contribute because $\rho^2 K\Theta K \equiv 0$ for $\rho > 0$ and produces a finite result at the limiting $\rho = 0$, or equivalently:

$$\lim_{\rho\to 0} \rho^3 K\Theta K = 0.$$

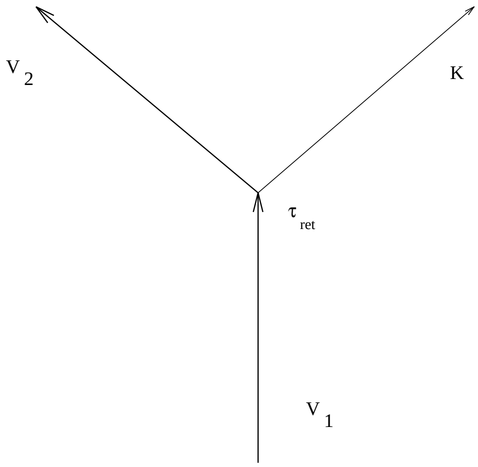

Figure 4. Classical picture of the fundamental quantum process: at $\tau_{\rm ret}$ an electron with a four-velocity V_1, changes it to V_2 by emitting a photon with a four-velocity K. $\tau_{\rm ret}$ is a singular point on the electron worldline because of this indeterminacy on its tangent. No infinity is ever involved.

10. The meaning of the classical field

A strict observation of the two geometric constraints ($R^2 = 0$ and $R\,{\rm d}R = 0$) in the LWS allows the introduction of the extended causality, the interpretation of these solutions in terms of creation and annihilation of particles, and the vision of the interacting electromagnetic field as composed of discrete point-like objects, the classical photons. The continuous picture of a wave and the idea of its continuous emission are just approximations valid for large-distance and macroscopic sources; it is justified for the normally large number of photons involved. The short-distance limit of CED is drastically changed with the extended causality: old inconsistencies, such as the non-integrability of the self-field energy tensor, disappears. The paradoxes associated to the electron equation of motion are all explained with the understanding that this is an effective equation, written in terms of averaged values and, therefore, limited on its applications and validity. The implicit non-locality of (70) is a consequence of the explicit bilocality of (1) and consequently of its energy tensor Θ: they both depend on $R = x - z(\tau_{\rm ret})$. Although $\varepsilon_1 \to 0$ in (70), the term $\langle \frac{2}{3}a^2V^\mu\rangle$ comes from the limiting $\varepsilon_2 \to \infty$. The fundamental (in the sense of irreducible) electromagnetic interaction is the exchange of a single photon. In classical physics this can be seen as the intersection of $A(x)$ with the light-cone generator that connects the two charges, as depicted in figure 3, and not by $A(x)$ itself, which rather represents the smearing of this interaction on the charge light-cone. That is why a space integration is necessary to retrieve the momentum carried out by the photon, which is the meaning of the RHS of equations (40) and (55). Their RHSs are point-by-point summations of the contribution of each light-cone generator. Equation (67) cannot therefore be a true equation of motion because it does not describe the single photon exchange, as required by the fundamental interaction, but an average of all photons inside the integration domain. This explains why (67) is not time-reversal invariant as a fundamental equation must be. The energy flux from the charge is, of course, non-zero at the point $z(\tau_{\rm ret})$, as it is detected at x, but it is indeed null at $z(\tau_{\rm ret} \pm {\rm d}\tau)$ as one concludes from (59) and (66). This has some noticeable consequences. It shows that the classical radiation process is discrete in time; this discreteness takes the ($\rho \to 0$)-limit to be revealed. At $\rho > 0$ this effect is masqueraded by the average character of $A(x)$. It is also in direct contradiction to Gauss's law, which makes no sense if the field is seen as the effect of a discrete exchange of particle-like objects, unless the field is taken as the average of these effects in space and time. It requires, therefore, a re-evaluation of the physical meaning of this law and the Faraday–Maxwell concept of fields. This question is also relevant to quantum theories (quantum mechanics and quantum field theory) because it deals with the reliability of the interaction description. How far does the classical field (that, in quantum field theory, one wants to quantize) really represents the experimentally observed interactions? This is closely related to the distinct contents of the Coulomb's and Gauss' law. While the first one gives a strict description of what is actually observed, i.e. a force between two charges, acting on each one along the straight line connecting them, the second one contains an extra assumption (the Faraday–Maxwell concept of field) that effectively *extends* this effect, observed at the charge position only, to all points in the space surrounding each one of the charges, regardless of the presence or not of the other.

The concept of a field existing everywhere around a single charge, regardless the presence of any other charge is an extrapolation of what is effectively observed. There is, therefore, a very deep distinction between Coulomb's law and Gauss's laws. This last one describes the *inferred* electric field as existing around a single charge, independent of the presence of the other charge. The electric field, as it is well known, is extracted from the Gauss's law through the integration of its flux across a closed surface, having the appropriate symmetry, *enclosing* the charge,

$$\boldsymbol{E}(x) = \hat{n}\frac{\int_V^x \rho\,\mathrm{d}v}{\int_{\partial V}\,\mathrm{d}S} \tag{94}$$

where $\hat{n}$ is the unit vector normal to the surface ∂V. Equation (94) adds evidence to the effective or average character of the Maxwell's concept of field; it also gives a hint to the meaning and origin of the field singularity. If the electric field can be visualized in terms of exchanged photons, then according to (94), the frequency or the number of these exchanged photons must be proportional to the enclosed net charge. If we take $\boldsymbol{E}$, as suggested by Gauss' law, as a measure of the average number of photons emitted/absorbed by a point charge, we can schematically write, $E \sim \frac{n}{4\pi r^2}$, where n is the number of photon per unit time crossing an spherical surface of radius r and centred on the charge. Then, the divergence of E when $r \to 0$ does not represent a physical fact such as an increasing number of photons, but just an increasing average number of photons per unit area, as the number of photons remains constant but the area tends to zero. So, a field singularity would have no physical meaning, it would just be a consequence of this average nature of the Maxwell's field.

In the modern perspective of seeing a fundamental electromagnetic interaction as the result of a single-photon exchange, the classical electromagnetic field describes rather the smearing of this interaction in time and in the space around each charge. No wonder one finds inconsistencies in the short-distance limit theory if one is replacing the interaction by its space average.

References

[1] Rohrlich F 1965 *Classical Charged Particles* (Reading, MA: Addison-Wesley)
[2] Jackson D 1975 *Classical Electrodynamics* 2nd edn (New York: Wiley) ch 4
[3] Parrot S 1987 *Relativistic Electrodynamics and Differential Geometry* (New York: Springer)
[4] Eliezer J C 1947 *Rev. Mod. Phys.* **19** 148
[5] Teitelboim C, Villaroel D and Van Weert Ch G 1980 *Rev. Nuovo Cimento* **3**
[6] Rowe E G P 1975 *Phys. Rev.* D **12** 1576
Rowe E G P 1978 *Phys. Rev.* D **18** 3639
Rowe E G P 1983 *Nuovo Cimento* B **73** 226
[7] Marx E 1975 *Int. J. Theor. Phys.* **14** 55–65
[8] Lozada A 1989 *J. Math. Phys.* **30** 1713

[9] Teitelboim C 1970 *Phys. Rev.* D **1** 1572

[10] de Souza M M 1995 The Lorentz–Dirac equation and the structure of spacetime *Preprint* UFES-DF-MMS.93/2, hep-th/9505169

[11] de Souza M M 1997 Discrete fields on the lightcone (in preparation)

[12] Synge J L 1975 *Ann. Mater. Pura Appl.* **84** 241

[13] Tabensky R 1976 *Phys. Rev.* D **13** 267