2

The Time Arrow of Radiation

After a stone has been dropped into a pond, one observes concentrically diverging ('defocusing') waves. Similarly, *after* an electric current has been switched on, one finds a retarded electromagnetic field that is coherently propagating *away* from its source. Since the fundamental laws of Nature, which describe these phenomena, are invariant under time reversal, they are equally compatible with the reverse phenomena, in which concentrically focusing waves (and whatever may be dynamically related to them – such as heat) would 'conspire' in order to eject a stone out of the water. Deviations of the deterministic laws from time reversal symmetry would modify this argument only in detail (see the Introduction). However, the reversed phenomena are never observed in Nature. In high-dimensional configuration space, the absence of dynamical correlations which would focus to create local effects characterizes the time arrow of thermodynamics (Chap. 3), or, when applied to wave functions, even that of quantum theory (see Sect. 4.3).

Electromagnetic radiation will here be considered to exemplify wave phenomena in general. It may be described in terms of the four-potential A^μ, which in the Lorenz gauge obeys the wave equation

$$-\partial^\nu \partial_\nu A^\mu(\boldsymbol{r}, t) = 4\pi j^\mu(\boldsymbol{r}, t) \ , \quad \text{with} \quad \partial^\nu \partial_\nu = -\partial_t^2 + \Delta \ , \tag{2.1}$$

using units with $c = 1$, the notations $\partial_\mu := \partial/\partial x^\mu$ and $\partial^\mu := g^{\mu\nu}\partial_\nu$, and Einstein's convention of summing over identical upper and lower indices. When an appropriate boundary condition is imposed, one may write A^μ as a functional of the sources j^μ. For two well known boundary conditions one obtains the *retarded* and the *advanced* potentials,

$$A^\mu_{\text{ret}}(\boldsymbol{r}, t) = \int \frac{j^\mu(\boldsymbol{r}, t - |\boldsymbol{r} - \boldsymbol{r}'|)}{|\boldsymbol{r} - \boldsymbol{r}'|} \mathrm{d}^3\boldsymbol{r}' \ , \tag{2.2a}$$

$$A^\mu_{\text{adv}}(\boldsymbol{r}, t) = \int \frac{j^\mu(\boldsymbol{r}, t + |\boldsymbol{r} - \boldsymbol{r}'|)}{|\boldsymbol{r} - \boldsymbol{r}'|} \mathrm{d}^3\boldsymbol{r}' \ . \tag{2.2b}$$

These two functionals of $j^\mu(\boldsymbol{r},t)$ are related to one another by a reversal of retardation time $|\boldsymbol{r}-\boldsymbol{r}'|$ – see also (2.5) and footnote 4 below. Their linear combinations are again solutions of the wave equation (2.1).

At this point, many textbooks argue somewhat mysteriously that 'for reasons of causality', or 'for physical reasons', only the retarded fields, derived from the potential (2.2a) according to $F^{\mu\nu}_{\mathrm{ret}} := \partial^\mu A^\nu_{\mathrm{ret}} - \partial^\nu A^\mu_{\mathrm{ret}}$, occur in Nature. This condition has therefore to be *added* to deterministic laws such as (2.1), which historically did indeed emerge from the asymmetric concept of causality. This example allows us to formulate in a preliminary way what seems to be meant by this *intuitive notion of causality*: correlated effects (that is, *nonlocal* regularities, such as coherent waves) must always possess a *local* common cause in their past.[1] However, this asymmetric notion of causality is a major *explanandum* of the physics of time asymmetry. As pointed out in the Introduction, it cannot be derived from the deterministic laws by themselves.

The popular argument that advanced fields are not found in Nature because they would require improbable initial correlations is known from statistical mechanics, but totally insufficient (see Chap. 3). The observed retarded phenomena are precisely as improbable among *all possible* ones, since they describe equally improbable *final* correlations. So their 'causal' explanation from an initial condition would beg the essential question.

Some authors take the view that retarded waves describe emission, advanced ones absorption. However, this claim ignores the fact that, for example, moving absorbers give rise to *retarded* shadows, that is, retarded waves which interfere destructively with incoming ones. In spite of the retardation, energy may thus flow from the electromagnetic field into an antenna. When incoming fields are present (as is generically the case), retardation does not necessarily mean emission of energy (see Sect. 2.1).

At the beginning of the last century, Ritz – following simular ideas by Planck and others – formulated a radical solution of the problem by postulating the exclusive existence of retarded waves *as a law*. Such time-directed *action at a distance* is equivalent to fixing the boundary conditions for the

[1] In the case of a *finite* number of local effects resulting from *one* local cause in the past, this situation is often viewed as a 'fork' in spacetime (see Horwich 1987, Sect. 4.8). However, this *fork of causality* should not be confused with the *fork of indeterminism* (in configuration space and time), which points to different (in general global) *potential states* rather than to different events (see also footnote 7 of Chap. 3 and Fig. 3.8). The fork of causality ('intuitive causality') may also characterize deterministic measurements and the documentation of their results, that is, the formation and distribution of information. It is related to Reichenbach's (1956) concept of *branch systems*, and to Price's (1996) *principle of independence of incoming influences* (PI^3). Insofar as it describes the cloning and spreading of information, it represents an *overdetermination* of the past (Lewis 1986), or the *consistency of documents*. It is these correlations which let the macroscopic past appear 'fixed', while complete documents about *microscopic* history would be in conflict with thermodynamics and quantum theory.

electromagnetic field in a universal manner. The field would then *not* describe any degrees of freedom on its own, but just describe retarded forces.

This proposal, a natural generalization of Newton's gravitational force, led to a famous controversy with Einstein, who favored the point of view that retardation of radiation can be explained by thermodynamical arguments. Einstein, too, argued here in terms of an action-at-a-distance theory (see Sect. 2.4). At the end of their dispute, the two authors published a short letter in order to state their different opinions. After an introductory sentence, according to which retarded and advanced fields are equivalent "in some situations", the letter reads as what appears to be also a verbal compromise (Einstein and Ritz 1909 – my translation):[2]

> While Einstein believes that one may restrict oneself to this case without essentially restricting the generality of the consideration, Ritz regards this restriction as not allowed in principle. If one accepts the latter point of view, experience requires one to regard the representation by means of the retarded potentials as the only possible one, provided one is inclined to assume that the fact of the irreversibility of radiation processes has to be present in the laws of Nature. Ritz considers the restriction to the form of the retarded potentials as one of the roots of the Second Law, while Einstein believes that the irreversibility is exclusively based on reasons of probability.

Ritz thus conjectured that the thermodynamical arrow of time might be explained by the retardation of electromagnetic forces because of the latter's universal importance for all matter. However, the retardation of hydrodynamical waves (such as sound) would then have to be explained quite differently – for example, by again referring to the thermodynamical time arrow.

A similar but less well known controversy had already occurred in the nineteenth century between Max Planck and Ludwig Boltzmann. The former, at that time still an opponent of statistical mechanics, understood radiation as a genuine irreversible process, while the latter maintained that the problem is not different from that in kinetic gas theory: a matter of improbable initial conditions (Boltzmann 1897). These different interpretations became relevant, in particular, in connection with the quantum hypothesis: are quanta *caused*

[2] The original text reads: "*Während Einstein glaubt, daß man sich auf diesen Fall beschränken könne, ohne die Allgemeinheit der Betrachtung wesentlich zu beschränken, betrachtet Ritz diese Beschränkung als eine prinzipiell nicht erlaubte. Stellt man sich auf diesen Standpunkt, so nötigt die Erfahrung dazu, die Darstellung mit Hilfe der retardierten Potentiale als die einzig mögliche zu betrachten, falls man der Ansicht zuneigt, daß die Tatsache der Nichtumkehrbarkeit der Strahlungsvorgänge bereits in den Grundgesetzen ihren Ausdruck zu finden habe. Ritz betrachtet die Einschränkung auf die Form der retardierten Potentiale als eine der Wurzeln des Zweiten Hauptsatzes, während Einstein glaubt, daß die Nichtumkehrbarkeit ausschließlich auf Wahrscheinlichkeitsgründen beruhe.*"

by the emission process (as Planck had believed – later called quantum jumps – see Sects. 4.3.6 and 4.5), or inherent to light itself?

In Maxwell's classical *field theory*, the problem does not appear as obvious as in action-at-a-distance theories, since every bounded region of spacetime may contain 'free fields', which possess neither past nor future sources *in this region*. Therefore, one can consistently understand Ritz's hypothesis only cosmologically: *all* fields must possess advanced sources ('causes') somewhere in the Universe. While the examples discussed above demonstrate that the time arrow of radiation cannot merely reflect *the way* boundary conditions are posed, the problem becomes even more pronounced with the time-reversed question: "Do all fields also possess a *retarded source* (a *sink* in time-directed terms) somewhere in the future Universe?" This assumption corresponds to the *absorber theory of radiation*, a T-symmetric action-at-a-distance theory to be discussed in Sect. 2.4. The observed asymmetries would then require an unusual cosmic time asymmetry in the distribution of such sources.

2.1 Retarded and Advanced Form of the Boundary Value Problem

In order to distinguish the indicated pseudo-problem that concerns only the definition of 'free' fields from the physically meaningful question, one has to investigate the general boundary value problem for hyperbolic differential equations (such as the wave equation). This can be done by means of Green's functions, defined as the solutions of the specific inhomogeneous wave equation with a point-like source:

$$-\partial^\nu \partial_\nu G(\boldsymbol{r}, t; \boldsymbol{r}', t') = 4\pi\delta^3(\boldsymbol{r} - \boldsymbol{r}')\delta(t - t') \ , \tag{2.3}$$

and an appropriate boundary condition in space and time. Some of the concepts and methods to be developed below will be applicable in a similar form in Sect. 3.2 to the Liouville equations (Hamilton's equations applied to ensembles of states of mechanical systems). Using (2.3), a solution of the general inhomogeneous wave equation (2.1) may then be written as a functional of its sources:

$$A^\mu(\boldsymbol{r}, t) = \int G(\boldsymbol{r}, t; \boldsymbol{r}', t') j^\mu(\boldsymbol{r}', t') \, \mathrm{d}^3 r' \, \mathrm{d}t' \ , \tag{2.4}$$

where the boundary condition for $G(\boldsymbol{r}, t; \boldsymbol{r}', t')$ determines that for $A^\mu(\boldsymbol{r}, t)$, too. Retarded or advanced solutions are obtained from Green's functions G_{ret} and G_{adv}, which are given by

$$G_{\substack{\mathrm{ret} \\ \mathrm{adv}}}(\boldsymbol{r}, t; \boldsymbol{r}', t') := \frac{\delta(t - t' \pm |\boldsymbol{r} - \boldsymbol{r}'|)}{|\boldsymbol{r} - \boldsymbol{r}'|} \ . \tag{2.5}$$

The potentials A^μ_{ret} and A^μ_{adv} resulting from (2.4) are thus functionals of sources only on the past *or* future light cones of their argument, respectively.

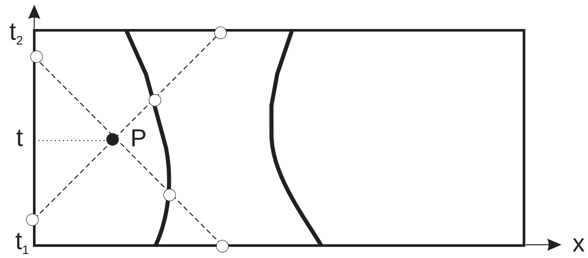

Fig. 2.1. Kirchhoff's boundary value problem, including initial, final and spatial boundaries. Sources (*thick world lines*) within the considered region and boundaries on both light cones (*dashed lines*) may in general contribute to the electromagnetic potential A^μ at the spacetime point P

By contrast, Kirchhoff's formulation of the boundary value problem allows one to express every specific solution $A^\mu(\boldsymbol{r}, t)$ of the wave equation by means of any Green's function $G(\boldsymbol{r}, t; \boldsymbol{r}', t')$. This can be achieved by using the three-dimensional Green theorem

$$\int_V \Big[G(\boldsymbol{r}, t; \boldsymbol{r}', t')\Delta' A^\mu(\boldsymbol{r}', t') - A^\mu(\boldsymbol{r}', t')\Delta' G(\boldsymbol{r}, t; \boldsymbol{r}', t')\Big]\, \mathrm{d}^3 r' \qquad (2.6)$$

$$= \int_{\partial V} \Big[G(\boldsymbol{r}, t; \boldsymbol{r}', t')\nabla' A^\mu(\boldsymbol{r}', t') - A^\mu(\boldsymbol{r}', t')\nabla' G(\boldsymbol{r}, t; \boldsymbol{r}', t')\Big] \cdot \mathrm{d}\boldsymbol{S}' ,$$

where $\Delta = \nabla^2$ is the Laplace operator, and ∂V is the boundary of the spatial volume V. Multiplying (2.3) by $A^\mu(\boldsymbol{r}', t')$, and integrating over $\boldsymbol{r}'$ and t' from t_1 to t_2 – on the right-hand side (RHS) by means of the δ-functions, while using the Green theorem and twice integrating by parts with respect to t' on the left-hand side (LHS), one obtains by further using (2.1):

$$A^\mu(\boldsymbol{r}, t) = \int_{t_1}^{t_2}\int_V G(\boldsymbol{r}, t; \boldsymbol{r}', t') j^\mu(\boldsymbol{r}', t')\, \mathrm{d}^3 r'\, \mathrm{d}t'$$

$$-\frac{1}{4\pi}\int_V \Big[G(\boldsymbol{r}, t; \boldsymbol{r}', t')\partial_{t'} A^\mu(\boldsymbol{r}', t') - A^\mu(\boldsymbol{r}', t')\partial_{t'} G(\boldsymbol{r}, t; \boldsymbol{r}', t')\Big]\, \mathrm{d}^3 r' \Big|_{t_1}^{t_2}$$

$$+\frac{1}{4\pi}\int_{t_1}^{t_2}\int_{\partial V} \Big[G(\boldsymbol{r}, t; \boldsymbol{r}', t')\nabla' A^\mu(\boldsymbol{r}', t') - A^\mu(\boldsymbol{r}', t')\nabla' G(\boldsymbol{r}, t; \boldsymbol{r}', t')\Big] \cdot \mathrm{d}\boldsymbol{S}'\, \mathrm{d}t'$$

$$\equiv \text{‘source term’} + \text{‘boundary terms’}. \qquad (2.7)$$

if the event P described by $\boldsymbol{r}$ and t lies within the spacetime boundaries. Here, both (past and future) light cones may contribute to the three terms occurring in (2.7), as indicated in Fig. 2.1.

The formal T-symmetry of this representation of the potential as a sum of a source term and boundary terms in the past *and* future can be broken by the choice of Green's functions. When using one of the two forms (2.5), the

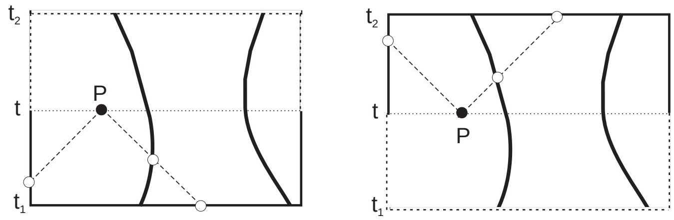

Fig. 2.2. Two representations of the same electromagnetic potential at time t by means of retarded or advanced Green's functions. They require data on partial boundaries (indicated by *solid lines*) corresponding to an initial or a final value problem, respectively

spacetime boundary required for determining the potential at time t assumes specific forms indicated in Fig. 2.2. Hence, the *same* potential can be written according to one or the other RHS of

$$\begin{aligned} A^\mu = \text{source term} + \text{boundary terms} &= A^\mu_{\text{ret}} + A^\mu_{\text{in}} \\ &= A^\mu_{\text{adv}} + A^\mu_{\text{out}} \ . \end{aligned} \tag{2.8}$$

For example, A^μ_{in} is here that solution of the *homogeneous* equations which coincides with A^μ for $t = t_1$. A^μ_{ret} and A^μ_{adv} vanish by definition for $t = t_1$ or $t = t_2$, respectively. *Any* field can therefore be described equivalently by an initial *or* a final value problem – with arbitrary boundary conditions. This result reflects the T-symmetry of the laws, while phenomenological causality is often used as an *ad hoc* argument for choosing G_{ret} rather than G_{adv}.

However, *two* free boundary conditions in the mixed form of Fig. 2.1 would in general not be consistent with one another, even if individually incomplete (see also Sects. 2.4 and 5.3). Retarded and advanced fields formally resulting from past and future sources, respectively, *do not add independently* (as sometimes assumed to describe a conjectured retro-causation) – they just contribute to different (or mixed) *representations* of the same field. In field theory, no (part of the) field 'belongs to' a certain source (in contrast to specific action-at-a-distance theories). Sources determine only the *difference* $A^\mu_{\text{out}} - A^\mu_{\text{in}}$ – similar to $T/i = S - 1$ in the interaction picture of the S-matrix. As can be seen from (2.8), this difference is identical to $A^\mu_{\text{ret}} - A^\mu_{\text{adv}}$. In causal *language*, where A^μ_{in} is regarded as given, the source 'creates' precisely its retarded field that has to be added to A^μ_{in} in the future of the source (where $A^\mu_{\text{adv}} = 0$).

Physically, *spatial* boundary conditions represent an interaction with the (often uncontrollable) spatial environment. For infinite spatial volume ($V = \mathbb{R}^3$), when the light cone cannot reach ∂V within finite time $t - t_1$, or in a closed universe, one loses this boundary term in (2.7), and thus obtains the *pure* initial value problem (for $t > t_1$),

$$A^\mu = A^\mu_{\mathrm{ret}} + A^\mu_{\mathrm{in}} \equiv \int_{t_1}^{t} \int_{\mathbb{R}^3} G_{\mathrm{ret}}(\boldsymbol{r},t;\boldsymbol{r}',t') j^\mu(\boldsymbol{r}',t')\,\mathrm{d}^3r'\,\mathrm{d}t' \tag{2.9}$$

$$+\frac{1}{4\pi}\int_{\mathbb{R}^3}\Big[G_{\mathrm{ret}}(\boldsymbol{r},t;\boldsymbol{r}',t_1)\partial_{t_1}A^\mu(\boldsymbol{r}',t_1) - A^\mu(\boldsymbol{r}',t_1)\partial_{t_1}G_{\mathrm{ret}}(\boldsymbol{r},t;\boldsymbol{r}',t_1)\Big]\,\mathrm{d}^3r'\ ,$$

and correspondingly the pure final value problem ($t < t_2$),

$$A^\mu = A^\mu_{\mathrm{adv}} + A^\mu_{\mathrm{out}} \equiv \int_{t}^{t_2} \int_{\mathbb{R}^3} G_{\mathrm{adv}}(\boldsymbol{r},t;\boldsymbol{r}',t') j^\mu(\boldsymbol{r}',t')\,\mathrm{d}^3r'\,\mathrm{d}t' \tag{2.10}$$

$$-\frac{1}{4\pi}\int_{\mathbb{R}^3}\Big[G_{\mathrm{adv}}(\boldsymbol{r},t;\boldsymbol{r}',t_2)\partial_{t_2}A^\mu(\boldsymbol{r}',t_2) - A^\mu(\boldsymbol{r}',t_2)\partial_{t_2}G_{\mathrm{adv}}(\boldsymbol{r},t;\boldsymbol{r}',t_2)\Big]\,\mathrm{d}^3r'\ .$$

The different signs at t_1 and t_2 are due to the fact that the gradient in the direction of the outward-pointing normal vector has now been written as a derivative with respect to t_1 (inward) or t_2 (outward).

So one finds precisely the retarded potential $A^\mu = A^\mu_{\mathrm{ret}}$ if $A^\mu_{\mathrm{in}} = 0$. (Only the 'Coulomb part', required by Gauß's law, must always be present by constraint. It can be regarded as the retarded *or* advanced consequence of the conserved charge.) In scattering theory, an initial condition fixing the incoming wave (usually described by a plane wave) is called a *Sommerfeld radiation condition*. Both conditions are to determine the actual situation. Therefore, the physical problem is not which of the two forms, (2.9) or (2.10), is *correct* (both are), but:

1. Why does the Sommerfeld radiation condition $A^\mu_{\mathrm{in}} = 0$ (in contrast to $A^\mu_{\mathrm{out}} = 0$) approximately apply in many situations?
2. Why are initial conditions more *useful* than final conditions?

The second question is related to the *historical nature* of the world. Answers to these questions will be discussed in Sect. 2.2.

The form (2.7) of the four-dimensional boundary value problem, characteristic of determinism in field theory, applies to partial differential equations of hyperbolic type (that is, with a Lorentzian signature $-{+}{+}{+}$). Elliptic type equations would instead lead to the Dirichlet or von Neumann problems, which require values of the field *or* its normal derivative, respectively, on a *closed* boundary (which in spacetime would have to include past *and* future). Only hyperbolic equations lead generally to 'propagating' solutions, which are compatible with free initial conditions. They are thus responsible for the concept of a *dynamical state* of the field, which facilitates the familiar concept of time.

The wave equation (with its hyperbolic signature) is known to be derivable from Newton's equations as the continuum limit of a spatial lattice of mass points, held at their positions by means of harmonic forces. For a linear chain, $m\mathrm{d}^2q_i/\mathrm{d}t^2 = -k\big[(q_i - q_{i-1} - a) - (q_{i+1} - q_i - a)\big]$ with $k > 0$, this is the limit $a \to 0$ for fixed ak and m/a. The crucial restriction to 'attractive' forces ($k > 0$) may here appear surprising, since Newton's equations are *always* deterministic, and allow one to pose initial conditions regardless of the type or sign of the forces. However, only bound (here oscillating) systems possess a *stable* position (here characterized by the lattice constant a). In the same limit, an elliptic differential equation (with signature ++++) would result for a lattice of variables q_i with repulsive forces ($k < 0$). This repulsion, though still representing deterministic dynamics, would cause the particle distances $q_i - q_{i-1}$ to explode immediately in the limit $k \to \infty$. The unstable solution $q_i - q_{i-1} = a$ is in this case the only eigensolution of the Dirichlet problem with eigenvalue 0 (derived from the *condition* of a bounded final state). Mathematically, the dynamically diverging solutions simply do not 'exist' any more in the continuum limit.

For second order wave equations, a hyperbolic signature forms the basis for all (exact or approximate) conservation laws, which give rise to the continuity of 'objects' in time (including the 'identity' of observers). For example, the free wave equation has solutions of a *conserved form* $f(z \pm ct)$, while the Klein–Gordon equation with a positive and variable 'squared mass' $m^2 = V(\boldsymbol{r}, t)$ has unitary solutions $\mathrm{i}\partial\phi(\boldsymbol{r}, t)/\partial t = \pm\sqrt{-\Delta + V}\phi(\boldsymbol{r}, t)$. This dynamical consequence of the spacetime metric, which leads to such 'wave tubes' (see also Sect. 6.2.1), is crucial for what appears as the inevitable 'progression of time' (in contrast to our freedom to move in space). However, the *direction* of this apparent flow of time requires *additional* conditions.

This section was restricted to the boundary value problem for fields in the presence of *given* sources. In reality, the charged sources depend in turn on the fields by means of the Lorentz force. The resulting coupled system of differential equations is still T-symmetric, while all consequences of the retardation regarding the *actual* electromagnetic fields, derived in this and the following section, remain valid. New problems will arise, though, from the *self-interaction* of point charges or elementary charged rigid objects (see Sect. 2.3).

2.2 Thermodynamical and Cosmological Properties of Absorbers

Wheeler and Feynman (1945, 1949) took up the Einstein–Ritz controversy about the relation between the two time arrows of radiation and thermodynamics. Their work essentially confirms Einstein's point of view, provided his 'reasons of probability' are replaced by 'thermodynamical reasons'. Statistical reasons by themselves are insufficient for deriving a thermodynamical arrow (see Chap. 3.) The major part of Wheeler and Feynman's arguments were again based on a T-symmetric action-at-a-distance theory, which is particularly well suited for presenting them in an historical context. From the point of view of local field theory (that is for good reasons preferred today), this picture may appear strange or even misleading. The description of their *absorber theory of radiation* will therefore be postponed until Sect. 2.4.