METHODOLOGICAL NOTES

Are the Kramers-Kronig relations for the dielectric permittivity of a material always valid?

D. A. Kirzhnits

P. N. Lebedev Physics Institute, USSR Academy of Sciences
Usp. Fiz. Nauk **119**, 357–369 (June 1976)

PACS numbers: 77.20.+y

CONTENTS

1. INTRODUCTION

In most texts on macroscopic electrodynamics the answer given to the question at the head of this article is an unqualified yes. However, the actual state of affairs turns out to be much more complicated. The correct answer to the question posed depends on precisely what quantity (the dielectric permittivity itself (DP, for short) or its inverse) and what range of wavelengths the Kramers–Kronig relations are written for (cf.[1–3]).

The importance of the question under discussion lies in the fact that the DP, being a concentrated source of information about the internal properties of a system and about the results of the action of different kinds of external probe on the system (cf.[1,4,5]), is one of the fundamental characteristics of systems with Coulomb interaction. The microscopic calculation of the DP is a problem of exceptional difficulty, yielding to solution only in the simplest cases. Therefore, an important role in the theory of condensed matter and plasmas is played by general restrictions on the value of the DP, which pertain to its analytic properties, its permissible limits of variation, etc., and are derived directly from general principles (causality, stability, positivity of the probability). The Kramers–Kronig relations, together with the consequences stemming from them, serve as the most important example of such restrictions.

Special attention is devoted in this article to one of these consequences, pertaining to the question of the allowed sign of the static DP (and, at the same time, to the question of the allowed sign of the static interaction between the electrons). The applications considered below, to the problem of radically raising the critical temperature of the superconducting transition, show that the question of the validity of the Kramers–Kronig relations goes far beyond the framework of the purely academic and has practical interest of considerable importance.

At the same time, the study of this question has definite methodological value, clearly demonstrating that for the proof of the validity of any relation between physical quantities an analysis of the actual realizability of the thought experiments lying at the basis of the definition of such quantities is necessary.

2. THE KRAMERS–KRONIG RELATIONS

We shall start from the definition of the DP $\varepsilon(\omega, \mathbf{k})$, meaning by this quantity the longitudinal DP of a uniform and isotropic system.[1)] We shall subject the latter to the action of an external source having the charge-density Fourier component $\delta\rho_e(\omega, \mathbf{k})$. As a result, there appears in the system the induced charge $\delta\rho(\omega, \mathbf{k})$, forming, together with $\delta\rho_e$, the total change $\delta\rho_t = \delta\rho_e + \delta\rho$ in the charge. The corresponding values of the displacement and field intensity are determined by the Maxwell equations

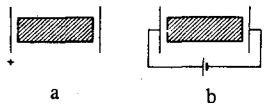
$$\operatorname{div} \mathbf{D} = 4\pi\delta\rho_e, \qquad \operatorname{div} \mathbf{E} = 4\pi\delta\rho_t.$$

In the language of the quantities introduced, the DP is defined by either of the relations

$$\begin{aligned} \delta\rho_e(\omega, \mathbf{k}) &= \varepsilon(\omega, \mathbf{k})\,\delta\rho_t(\omega, \mathbf{k}), \\ \mathbf{D}(\omega, \mathbf{k}) &= \varepsilon(\omega, \mathbf{k})\,\mathbf{E}(\omega, \mathbf{k}). \end{aligned} \tag{1}$$

We shall make precise the concept of the external

[1)] Incidentally, a number of remarks pertaining to the transverse DP and also to crystalline media will be made below.



a b

FIG. 1.

source, to which belong, by definition, the charges that do not appear in the composition of the system under consideration. These charges are, primarily, those positioned outside the system (e.g., the charges on the capacitor plates between which the system itself is situated; see Fig. 1). By such means we can produce a $\delta\rho_e$ or D only with $k \sim L^{-1} \to 0$, where L is the macroscopically large size of the system. Therefore, external sources of this type are associated only with the long-wavelength limit $\varepsilon(\omega, 0)$ of the DP.

To realize the case $\mathbf{k} \neq 0$ it is necessary to place the external sources inside the system itself ($\delta\rho_e \sim \exp(i\mathbf{k}\cdot\mathbf{x})$); these are often also called extraneous sources. In this case it is necessary to fix the state of these sources, by assigning to them, e.g., a large mass and thereby eliminating the back-action of the system itself on them. Otherwise, we would lose the possibility of distinguishing the internal and external charges and the situation would reduce, in essence, to treating a new, more complicated system.

The dielectric permittivity describes the reaction (response) of the system to an external perturbation. The quantity establishing the relation between the action on the system and its result is called the response function:

$$\text{(result of perturbation)} = \text{(response function)} \times \text{(perturbation)}. \qquad (2)$$

This relationship has a causal character ("cause always precedes effect in time"), the role of the effect being played by the left-hand side of (2) and that of the cause by the second factor of the right-hand side. It follows from this (cf., e.g., [1]) that the response function necessarily satisfies a relation of the Kramers–Kronig type. Therefore, the purpose of the following discussions will be to elucidate the question of which of the quantities (ε or $1/\varepsilon$), and under what conditions, can be regarded as the response function.

The meaning of the concept of the external perturbation appearing in Eq. (2) also needs to be made precise. By definition, this is what we call a perturbation which does not depend on the state of the acted-upon system and which can be varied, switched on and switched off at will; an event-cause (for more detail, cf. [6]) necessarily possesses just such properties. It is of importance that by far not every perturbation realizable by an external source satisfies this definition.

A perturbation arising from an external source of the second of the types considered above (a source located inside the system and free from the back-influence of the system) always falls under this definition. In this case, the perturbation is related to the quantities $\delta\rho_e$ and D and its result to $\delta\rho_t$ and E. Comparing (1) with (2), we see that the quantity $1/\varepsilon(\omega, \mathbf{k})$ can be regarded as the response function, under all conditions. Correspondingly, for all[2)] values of $\mathbf{k}$ the Kramers–Kronig relation for the inverse DP will be valid[1]:

$$\varepsilon^{-1}(\omega, \mathbf{k}) = 1 + \frac{1}{\pi}\int_0^\infty \frac{d\omega'^2 \operatorname{Im}\varepsilon^{-1}(\omega', \mathbf{k})}{\omega'^2 - \omega^2 - i\delta}. \qquad (3)$$

It is essential that, in these discussions, we cannot interchange E and D and, correspondingly, regard the DP itself as the response function. The point is that we cannot regard the intensity as the action (it depends on the state of the system) and the displacement as a characteristic of the result of the action (on the contrary, it does not depend on the state of the system).

As regards an external source situated outside the system, it gives rise to an external perturbation only under the condition that the plates of the capacitor are disconnected and we can control the magnitude of the charge on them (see Fig. 1a). In this case, as before, we can regard the response function to be the quantity $1/\varepsilon$, but with $k \to 0$. But if the plates are connected through a battery and we control the magnitude of the potential difference across the capacitor (see Fig. 1b), the corresponding external perturbation is not related to the external source (the charge on the plates). In the conditions under consideration, the latter itself depends on the state of the system, flowing toward the plates or away from them as this state changes. The external perturbation is now determined by the magnitude of the total potential difference across the plates, i.e., by the field intensity E, and the result of this perturbation by the magnitude of the external charge or of the displacement D. Accordingly, besides $1/\varepsilon(\omega, \mathbf{k})$ we can regard the quantity $\varepsilon(\omega, 0)$ as a response function and the Kramers–Kronig relations for the DP itself are found to be valid, but only in the long-wavelength limit:

$$\varepsilon(\omega, 0) = 1 + \frac{1}{\pi}\int_0^\infty \frac{d\omega'^2 \operatorname{Im}\varepsilon(\omega', 0)}{\omega'^2 - \omega^2 - i\delta}. \qquad (4)$$

The examples considered exhaust the possible types of action on a system that can be realized by means of a longitudinal field. As regards the response to a transverse field, for $k = 0$ for the transverse DP $\varepsilon_t(\omega, \mathbf{k})$, which coincides in this limit with the longitudinal DP, both relations (3) and (4) are valid. For $\mathbf{k} \neq 0$ the Kramers–Kronig relation for the quantity $[\varepsilon_t(\omega, \mathbf{k}) - (k^2c^2/\omega^2)]^{-1}$ holds.[3)]

We note that to derive (3) we can use instead of the principle of causality the so-called "spectrality" condition, relating the quantity $1/\varepsilon$ to the correlation of the fluctuations $\langle E_i E_j \rangle$,[5] introducing a complete intermediate set of functions, and taking into account that the

---

[2)] Contrary to widespread opinion, the DP, like the first relation of (1), has an exact microscopic meaning and can also be used for large values of k, when the concept of the intensity as an average field becomes meaningless (cf.[4]).

[3)] We point out that stronger statements pertaining to the transverse DP, which, however, are insufficiently substantiated, are contained in[2].

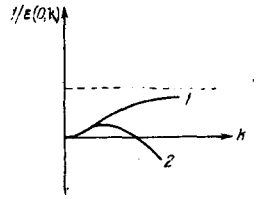


FIG. 2.

frequencies corresponding to them are real. It is in precisely this way that the Kallen–Lehmann relations for the vacuum photon Green function in quantum field theory, which serve as a direct analog of the relation (3), are derived. We emphasize that by virtue of the relativistic invariance the Green function depends on the single combination $\omega^2 - c^2k^2$. For this reason the Kallen-Lehmann relations are also valid for the inverse Green function (the analog of the quantity ε) for all values of its argument.[7]

3. CONSEQUENCES OF THE KRAMERS-KRONIG RELATIONS

The most important consequence, with which we shall be solely concerned below, pertains to the permissible limits of variation of the static DP $\varepsilon(0, \mathbf{k})$. We shall start from the relation

$$\operatorname{Im} \varepsilon^{-1}(\omega, \mathbf{k}) = -\frac{4\pi^2 e^2}{k^2} F(\omega, \mathbf{k}), \tag{5}$$

where $F > 0$ is the form factor for inelastic scattering of electrons by the system considered and determines the probability that the electron loses energy ω and momentum $\mathbf{k}$.[1] Hence, $\operatorname{Im}\varepsilon^{-1} < 0$, $\operatorname{Im}\varepsilon > 0$, and the relations (3) and (4) for $\omega = 0$ give the following restrictions: the relation (3) leads to the inequality $1/\varepsilon(0, \mathbf{k}) < 1$ or, which is the same thing, to one of the inequalities

$$\varepsilon(0, \mathbf{k}) > 1, \qquad \varepsilon(0, \mathbf{k}) < 0, \tag{6}$$

while the relation (4) leads to the inequality

$$\varepsilon_0 \equiv \lim_{k \to 0} \varepsilon(0, \mathbf{k}) > 1. \tag{7}$$

As can be seen from the results, the absence of Kramers–Kronig relations for the DP itself for $k \neq 0$ opens up the possibility of negative values for the static DP. This possibility was discussed in[1], but was rejected by unconvincing qualitative arguments. Below we shall consider these and other arguments in favor of the necessity of a positive sign of $\varepsilon(0, \mathbf{k})$ and show them to be incorrect.

A negative sign of the static DP implies "over-screening" of the external charge, as a result of which its field acquires a sign opposite to that of the field in the vacuum. Of more importance is another consequence: the effective interaction between the particles also has a sign opposite to that in the vacuum. In fact, we shall consider the energy of the interaction of a static external charge $\delta\rho_e(0, \mathbf{k})$, producing a field $4\pi\delta\rho_t(0, \mathbf{k})/k^2$ inside the system, with itself (cf. (1)):

$$\frac{1}{2}\sum_{\mathbf{k}} \frac{4\pi}{k^2} \delta\rho_e^*(0, \mathbf{k})\, \delta\rho_t(0, \mathbf{k}) = \frac{1}{2}\sum_{\mathbf{k}} \frac{4\pi}{k^2 \varepsilon(0, \mathbf{k})} |\delta\rho_e(0, \mathbf{k})|^2.$$

From this it can be seen that the Fourier transform of the static interaction of two particles with charge e has the form[4)]

$$V(\mathbf{k}) = \frac{4\pi e^2}{k^2 \varepsilon(0, \mathbf{k})}. \tag{8}$$

This relation serves as the obvious generalization of the well-known elementary formula for Coulomb's law in a medium. Below we shall return again to the discussion of formula (8) and its consequences.

To conclude this section we shall consider the question of the behavior of the quantity $\varepsilon(0, \mathbf{k})$ with variation of k (cf. (6), (7)). For $k = 0$ (and also, from continuity considerations, in a region of relatively small k that is wider the smaller is the role of spatial dispersion in the system), this quantity is positive and greater than unity. With increase of k, either these conditions continue to be fulfilled or $\varepsilon(0, \mathbf{k})$ becomes infinite, changing its sign. More adequate is the language based on the quantity $1/\varepsilon$: in the first case it always lies between zero and unity (Fig. 2, curve 1) and in the second it passes through zero (Fig. 2, curve 2).

We note in this connection that the derivation of the inequality $\varepsilon > 1$ given in[8] (Sec. 14) actually permits negative values of ε, as can be seen after going over to the $1/\varepsilon$ language. This derivation is based on the necessity of a positive sign of the change of ε (or, which is the same thing, a negative sign of the change of $1/\varepsilon$) when the initial value is $\varepsilon = 1$. But this in no way excludes the possibility that, as it decreases, the function $1/\varepsilon$ will pass through zero, becoming negative. In the language of ε this possibility is masked by the inevitable divergence of ε to infinity.

4. CRITERIA FOR THE STABILITY OF A SYSTEM

In this section it will be shown that the inequalities (6) and (7) obtained above are none other than criteria for the stability of the system. Simultaneously, those arguments against the second inequality (6) based on considerations of stability will be overthrown.

We shall formulate the stability problem in a general form, confining ourselves for simplicity to the case of zero temperature.[5)] Suppose that we are interested in the stability of the system with respect to variations of a certain physical quantity—a "field" $\varphi(\mathbf{x})$; we shall as-

---

[4)] Referring to the book[3] for the details, we stress that (8) describes the interaction of just the external charges, the state of which is fixed. The law of interaction of internal charges of the system, which experience the back-action of the sytem, differs, generally speaking, from (8). However, in many important cases, this difference is insignificant.

[5)] With regard to this formulation of the problem, cf.[9,10]. A modern formulation of the stability problem (the effective-potential method), developed in recent years in connection with the study of spontaneous symmetry breaking in quantum field theory and statistics, is contained in, e.g., the papers[11].

sume a "current" $j(\mathbf{x})$ to be conjugate with respect to the field. We shall formulate the problem of seeking a functional $E\{\varphi\}$ which would depend on the arbitrary (nonequilibrium) function φ and which would have a minimum with respect to φ in the equilibrium state. The function φ that realizes the minimum defines the equilibrium field, and the actual conditions for the minimum play the role of criteria of the stability of the system.

In order to make a state of the system with arbitrary φ an equilibrium state, it is necessary to introduce a specially chosen external current j by adding a term $\hat{\mathcal{H}}' = \int d\mathbf{x} j\hat{\varphi}$ to the Hamiltonian $\hat{\mathcal{H}}$ of the system. Denoting $E\{j\} = \langle \hat{\mathcal{H}} + \hat{\mathcal{H}}' \rangle$, where the brackets denote averaging over the state with the current, we have

$$\frac{\delta E\{j\}}{\delta j(\mathbf{x})} = \langle \hat{\varphi}(\mathbf{x}) \rangle \equiv \varphi(\mathbf{x}). \tag{9}$$

The functional $E\{j\}$ has an intermediate character and has no minimization properties with respect to j (this is already clear from (9) for $\varphi \neq 0$). In order to go over to the required functional $E\{\varphi\}$ it is necessary to subtract from $E\{j\}$ the "superfluous" work associated with the external current and change from j to the argument φ, using (9). This reduces to a Legnedre transformation

$$E\{\varphi\} = E\{j\} - \int d\mathbf{x} j\varphi = \langle \mathcal{H} \rangle,$$

in which the current appearing in the averaging must be expressed in terms of φ. The conditions for the minimum of the expression obtained, which has the meaning of the energy of the system for arbitrary φ, take the form

$$\frac{\delta E\{\varphi\}}{\delta \varphi(x)} = -j(\mathbf{x}) = 0, \quad \frac{\delta^2 E\{\varphi\}}{\delta\varphi(\mathbf{x})\,\delta\varphi(\mathbf{x}')} = -\frac{\delta j(\mathbf{x}')}{\delta\varphi(\mathbf{x})} > 0. \tag{10}$$

We shall show now that the conditions (6) coincide with the criteria for stability of the system against the spontaneous appearance of density waves in it (see the next section); this corresponds to the choice $\varphi = \rho$, $j = U_e$, where U_e is the potential of the external field. Then the first condition (10) gives $U_e = 0$ and the second leads to the inequality

$$\frac{\delta^2 E\{\rho\}}{\delta\rho(0, \mathbf{k})\,\delta\rho(0, -\mathbf{k})} = -\frac{\delta U_e(0, \mathbf{k})}{\delta\rho(0, \mathbf{k})} = -\frac{1}{\chi(0, \mathbf{k})} > 0$$

(equilibrium static density waves are considered). Here we have introduced the susceptibility

$$\chi(\omega, \mathbf{k}) = \frac{\delta\rho(\omega, \mathbf{k})}{\delta U_e(\omega, \mathbf{k})} = \frac{k^2}{4\pi}\left(\frac{1}{\varepsilon(\omega, \mathbf{k})} - 1\right).$$

It is obvious that the criterion obtained coincides exactly with the condition (6).

This is not surprising, since the spontaneous appearance of density waves should necessarily accompany an unstable state in the excitation spectrum of a spatially uniform system (a zero of the function $\varepsilon(\omega, \mathbf{k})$ with $\omega^2 < 0$), and this is forbidden by the Kramers–Kronig relation (3). At the same time, the violation of the Kramers–Kronig relation for the DP itself for $k \neq 0$ would mean the appearance of poles of $\varepsilon(\omega, \mathbf{k})$ with $\omega^2 < 0$, which do not have a direct physical meaning. On the other hand, the instability of the system leads to direct violation of the causality principle because of the impossibility of distinguishing a density wave arising as a response of the system in advance of the external perturbation and one arising as a result of the instability of the system. We note that the method formulated above coincides in the problem under consideration with the well-known variational principle of quantum many-body theory (in recent years it is most often called the density-functional method; cf., e.g., [12]).

The analysis carried out corresponds to the case $k \neq 0$ (density waves whose appearance does not violate the normalization to a fixed volume). The case $k = 0$ is special and corresponds to study of the stability with respect to a spontaneous change in the average density of the system (collapse of the system). Stability requires that the value of the bulk modulus or of the square of the sound velocity s be positive. The relation[1] $s^2 = \omega_p^2/k^2(\varepsilon_0 - 1)$, where ω_p is the plasma frequency of the electrons, leads to the inequality (7).

5. STABILITY IN THE LANGUAGE OF MACROSCOPIC ELECTRODYNAMICS

It remains for us to consider critically the arguments, based on stability considerations,[1,13] against the possibility of a negative sign of the static DP. These arguments are based on expressions for the field energy of the system,

$$\frac{1}{8\pi}\sum_{\mathbf{k}}\frac{|\mathbf{D}(0, \mathbf{k})|^2}{\varepsilon(0, \mathbf{k})} = \frac{1}{8\pi}\sum_{\mathbf{k}}\varepsilon(0, \mathbf{k})|\mathbf{E}(0, \mathbf{k})|^2, \tag{11}$$

which indeed have no minimum for $\varepsilon(0, \mathbf{k}) < 0$.

The left-hand side of the expression (11), having the meaning of the energy of infinitely heavy ions producing an external (with respect to the electrons) field D, is considered in[1]. The appearance of a wave of D, i.e., the rearrangement of the ionic subsystem, is regarded as a manifestation of instability. In[13] the right-hand side of the expression (11) is taken as fundamental and the appearance of a field-intensity wave is regarded as manifesting instability when the sign of the static DP is negative.

In fact, both these conclusions are incorrect, since the functionals (11) need not have a minimum at all in the equilibrium state. Below we shall give the correct solution of the problem of the stability of the system, in the language of the electrodynamic parameters D and E, basing the discussion on the general method of Sec. 4. As the result of the treatment we again arrive at the inequalities (6) and (7).

We begin by studying the stability with respect to the appearance of a field E; accordingly, $\varphi = \mathbf{E}$. In this case we are concerned with a connected system, corresponding to Fig. 1a. The change in the energy of the system as a result of the action of the external current (the role of which is played by the charges on the plates)

is determined by the well known expression $E\{j\} = (1/8\pi)\sum_k |D|^2/\varepsilon$, whence $j = D/4\pi$ and $E = D/\varepsilon$. Performing a Legendre transformation and changing to the variable E, we obtain $E\{\varphi\} = -(1/8\pi)\sum_k \varepsilon |E|^2$. In this expression it is necessary to retain only that part of it which pertains to the actual system under consideration. Taking into account that the external charges (i.e., the displacement D) are fixed in our thought experiment and separating out the corresponding constant term, we arrive at the final expression for that functional of the intensity E which should indeed have a minimum with respect to E, and which differs radically from the right-hand side of the expression (11):

$$E\{E\} = \frac{1}{8\pi}\sum_k \varepsilon(0, k)[\varepsilon(0, k) - 1]|E(0, k)|^2. \qquad (12)$$

The stability condition stemming from this coincides exactly with (6).

The study of the stability with respect to the spontaneous appearance of a displacement D, corresponding to the scheme of Fig. 1b (disconnected system),[6] is carried out analogously. In this case, $\varphi = D$, $E\{j\} = -(1.8\pi)\sum_k \varepsilon |E|^2$ (cf.[8]), $j = -E/4\pi$ and $D = \varepsilon E$. Performing the same operations as above, we find $E\{\varphi\} = (1/8\pi)\sum_k |D|^2/\varepsilon$. This expression formally coincides with the left-hand side of (11). However, it has a completely different meaning, as it corresponds to a physically different formulation of the problem; in particular, this invalidates the words about the rearrangement of the ionic subsystem (see above), which is already impossible by virtue of the infinitely large mass of the ions. Of more importance is another point—we can use only the long-wavelength part of the expression obtained. This follows from the large size of the system (cf. Sec. 2), and also from the argument that, by virtue of the equation div $D = 4\pi\rho_e$, in the absence of external sources only a spatially uniform displacement can arise spontaneously inside the system. Taking this circumstance into account and subtracting, as above, the part pertaining to the external sources themselves (the potential difference across the plates (i.e., the intensity E) is now fixed), we obtain the final expressions:

$$E\{D\} = \frac{1}{8\pi}\frac{\varepsilon_0 - 1}{\varepsilon_0^2}|D(0, 0)|^2. \qquad (13)$$

The inequality (7) is a direct consequence of this.

To conclude this section we note that the functionals obtained ((12 and 13)) have a direct physical meaning, inasmuch as each of the states competing in the calculation of the minimum can be realized in practice by a suitable choice of the external current. In this they differ from the functionals introduced in the book[8] (Sec. 18), which have a formal meaning and do not always possess a minimum at the equilibrium state.

6. COHERENT CRYSTALLIZATION

The outcome of the treatment carried through in the last two sections is that the conditions (6) and (7) play the role of conditions for the stability of the system. When they are violated the system should inevitably rearrange because of the instability of its previous state. If the condition (7) is violated, then, as already pointed out, collapse of the system will occur: it begins to contract irresistibly, increasing its density without limit. A more complicated and interesting pattern arises when the condition (6) is violated. We have already mentioned above the appearance in this case of density waves, implying the violation of the translational invariance of the system under consideration. In this section we shall consider in more detail the character of the state which arises in such a rearrangement; this state has previously been called a "coherent crystal".[14][7]

Such an analysis in the context of this article is all the more justified by the fact that the following question can arise: does the above investigation of the stability of an idealized uniform system against its transition to a nonuniform state have any meaning at all if we know for certain that such a system is unstable (at low temperatures) with respect to the formation of an ordinary crystalline lattice? In fact, the answer to this question turns out to be positive, since the properties of an ordinary and a coherent crystal are radically different in many respects and the spontaneous appearance of such structures corresponds to two independent modes of instability of the uniform state of the system.

Referring to[14] for the details, we shall list briefly the principal features common to and distinguishing the structures of ordinary and coherent crystals:

a) Both types of structure are characterized by violation of translational symmetry—the appearance of a spatially periodic distribution of the particle density.

b) Both types of structure possess shear rigidity and belong to the category of solids.

c) The period of an ordinary crystal is determined by the concentration of the particles and that of a coherent crystal by the law of interaction of the particles. In the latter case the reciprocal-lattice vector belongs to the range of k in which the conditions (6) are violated, i.e., the quantity $1/\varepsilon(0, k)$ is greater than unity.

d) The particles of an ordinary crystal are localized near the sites of the crystal lattice and the particles of a coherent crystal move relatively freely over its whole volume.

e) An ordinary crystal is a structure arising as a result of many-particle correlations in the system. A coherent crystal, on the other hand, can also be considered in the framework of a single-particle picture: the periodic distributions of the density and of the self-consistent field support each other. This corresponds

[6] In general, a quantity which, in the language of response functions, corresponds not to the "perturbation" but to its "result" (cf. Sec. 2) can arise spontaneously.

[7] The pioneering papers in this field are due to A.A. Vlasov and also to A. Overhauser (for a detailed bibliography, see in[14]).

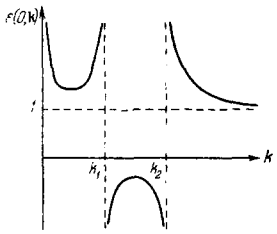


FIG. 3.

to a coherent density wave due to the Bose condensation of "particle-hole" pairs (for Fermi systems) or of the bosons themselves (for Bose systems) in a state with definite wave vector $\mathbf{k} \neq 0$ and definite phase.

We shall not touch upon other differences between ordinary and coherent crystals (see[14]), which are, in a certain sense, opposite types of periodic solid-state structures. The widely studied "quantum crystals"[15] should be assigned to an intermediate type, for which properties of both of the crystal types compared are characteristic.

7. THE SIMPLEST MODELS

In this section we given an illustration of the material described, using the example of the simplest theoretical models. The general expression for the DP can be written in the form

$$\varepsilon(\omega, \mathbf{k}) = 1 + 4\pi \sum_i \alpha_i(\omega, \mathbf{k}), \tag{14}$$

where the summation runs over the different subsystems of the substance (the conduction electrons, ions, etc.), and α_i is the polarizability of a subsystem, connected by the relation $\Pi = -k^2\alpha$ with its most important microscopic characteristic—the polarization operator Π.

The polarizability of the conduction electrons ($i = 1$) has the following well-known limiting expressions:

$$\begin{aligned} \alpha_1(\omega, \mathbf{k}) &= \frac{\varkappa^2}{4\pi k^2} \quad \left(\frac{\omega\varkappa}{\omega_p} \ll k \ll \varkappa\right), \\ \alpha_1(\omega, \mathbf{k}) &= -\frac{\omega_p^2}{4\pi\omega^2} \quad \left(k \ll \frac{\omega\varkappa}{\omega_p}\right), \end{aligned} \tag{15}$$

where $\varkappa$ is the inverse Debye screening length. The polarizability of the ion subsystem ($i = 2$) in the simplest "jellium" model in which the ions are regarded as a continuum of free particles has the form

$$\alpha_2(\omega, \mathbf{k}) = -\frac{\omega_{pi}^2}{4\pi\omega^2}, \tag{16}$$

where ω_{pi} is the plasma frequency of the ions. We note that, even in the framework of this very simple model (which, incidentally, is not suitable in the region of low frequencies), the static DP is negative (it tends to $-\infty$). In the more complicated model in which the ions are assumed to be bound and to have the spectrum of the "bare" frequencies $\omega_0(\mathbf{k})$,

$$\alpha_2(\omega, \mathbf{k}) = -\frac{\omega_{pi}^2}{4\pi(\omega^2 - \omega_0^2(\mathbf{k}) + i\delta)}. \tag{17}$$

The latter model is already able, in principle, to describe a system in which $\varepsilon(0, \mathbf{k}) < 0$. Starting from the expression

$$\varepsilon(0, \mathbf{k}) = 1 + 4\pi\alpha_1(0, \mathbf{k}) + \frac{\omega_{pi}^2}{\omega_0^2(\mathbf{k})},$$

we see that for this it is necessary in any case that the condition $\omega_0^2(\mathbf{k}) < 0$ be fulfilled. In other words, instability of the "bare" ionic subsystem existing before its interaction with the conduction electrons is switched on is necessary. However, the final physical phonon spectrum will be stable when the conditions (6) and (7) are fulfilled. We shall give the simplest mathematical model for $\omega_0^2(\mathbf{k})$, satisfying all the necessary requirements:

$$\omega_0^2(\mathbf{k}) = A\omega_{pi}^2 \frac{(k^2 - k_1^2)(k^2 - k_2^2)}{(k_1^2 - k_2^2)^2}, \quad A < 4.$$

Figure 3 corresponds to this. However, the question of the physical realization of such a model remains open as yet.[8)]

If systems with $\varepsilon(0, \mathbf{k}) < 0$ really exist, they can be transformed to the coherent-crystal state (cf. Sec. 6) by adding oscillators of frequency $\Omega \gg \omega_{pi}$, uniformly distributed over the volume, which would have produced a DP $\bar{\varepsilon} > 1$ *in vacuo* (a generalized "jellium" model). The corresponding polarizability ($i = 3$) has the form

$$\alpha_3(\omega, \mathbf{k}) = -\frac{(\bar{\varepsilon} - 1)\Omega^2}{4\pi(\omega^2 - \Omega^2 + i\delta)}.$$

It increases the original DP of the system by the amount $\bar{\varepsilon} - 1$. It is clear for $\varepsilon(0, \mathbf{k}) < 0$ a suitable choice of $\bar{\varepsilon}$ makes it possible to carry the system over into the interval $0 < \varepsilon(0, \mathbf{k}) < 1$ forbidden by conditions (6), and this leads to the formation of a coherent crystal.

To conclude this section we shall discuss a question that often arises when we compare the Kramers–Kronig relations for ε and $1/\varepsilon$, which express the fact that these response functions are retarded. At first sight it seems that these properties are incompatible (simultaneously, E is retarded with respect to D and, conversely, D is retarded with respect to E). In fact, in the nonrelativistic theory that we are considering the structures of the time-dependence of E and D are entirely similar, irrespective of which of these quantities describes the "perturbation" and which its "result": both these quantities are nonzero only at times after the "perturbation" is switched on.

It is easy to convince oneself of this using the simple model (17), omitting the electronic polarizability (16) for simplicity:

[8)] In this connection we point out the stability condition $\varepsilon(0, \mathbf{k})[1 - (g_0^2 k^2/4\pi e^2)] > 0$ (cf. Chap. 3 of[3], and[16]), where g_0 is the "bare" electron-phonon coupling constant; for $g_0^2 k^2 > 4\pi e^2$, negative values of $\varepsilon(0, \mathbf{k})$ are possible. We note that allowance for Umklapp processes in the crystal lattice can also lead to the expression (17) with $\omega_0^2 < 0$ (cf.[13,17]).

$$\varepsilon(\omega, \mathbf{k}) = 1 + \frac{\omega_{pi}^2}{\omega_0^2 - \omega^2 - i\delta}. \quad \varepsilon^{-1}(\omega, \mathbf{k}) = 1 - \frac{\omega_{pi}^2}{\omega_0^2 + \omega_{pi}^2 - \omega^2 - i\delta}. \qquad (18)$$

These two quantities have the same rules for going round the singularities[9)] and, therefore, in the t-representation, both have the "retardation" factor $\theta(t) = 1(t > 0)$ or 0 $(t < 0)$:

$$\varepsilon(t, \mathbf{k}) = \delta(t) + \theta(t)\frac{\omega_{pi}^2}{\omega_0}\sin(\omega_0 t),$$

$$\varepsilon^{-1}(t, \mathbf{k}) = \delta(t) - \theta(t)\frac{\omega_{pi}^2}{\sqrt{\omega_0^2 + \omega_{pi}^2}}\sin(\sqrt{\omega_0^2 + \omega_{pi}^2}\, t).$$

The same properties are possessed by the fields $\mathbf{E}(t)$ and $\mathbf{D}(t)$ themselves.

8. APPLICATIONS TO THE THEORY OF SUPERCONDUCTIVITY

As already noted in Sec. 3, from the results of this article there follows, in particular, the following statement: systems within which the electrons do not repel each other, as they do in the vacuum, but attract each other (at large values of k, i.e., at comparatively short distances) can, in principle, exist. This statement, as we shall show below, can turn out to be important for the physics of the phenomenon of superconductivity, since the character of the static interaction of the electrons at relatively large values of k is important for the formation of Cooper pairs and for the actual appearance of this phenomenon.

The statement made often seems trvial to those people who are familiar with the theory of superconductivity from its popular accounts, in which the formation of the Cooper pairs is explained by precisely the electron-electron attraction induced by their interaction with the lattice (phonons).

Actually, however, it has been well known for some time[3,18] that the electron-electron interaction V' which is responsible for their "pairing" and which, in a superconductor, should in fact have an attractive character, differs appreciably from the true electron-electron interaction V (cf. (8)). It turns out that if we write the latter in the form $V = V_{ph} + V_c$, where V_{ph} is the phonon part (the attraction) and V_c is the Coulomb part (the repulsion), then $V' = V_{ph} + \alpha V_c$, where α is less than unity. Therefore, from the condition $V' < 0$, necessary for "pairing," it follows that the true interaction V is smaller than the quantity $(1 - \alpha)V_c$ and can be either negative or positive. Moreover, the available experimental data support the view that forces of repulsion between the particles dominate in the known superconductors.

At the level in which we are interested, the most important consequence of the statement discussed concerns the existence of an upper bound for the critical temperature T_c of the superconducting transition. If the Kramers–Kronig relation (4) were valid for large values of k and, consequently, the restriction $\varepsilon(0, \mathbf{k}) > 1$ were fulfilled, then the inequality $V > 0$ would, in any case, follow. At the same time, the "pairing" interaction would turn out to be bounded $(|V'| < (1 - \alpha)V_c)$, and with it the magnitude of T_c.[13,17] An estimate performed in the first of these papers gives for the upper bound of T_c a value close to that already attained for the record-breaking superconductors, and if this estimate is reliable it would leave little hope of a radical raising of the critical temperature. Although this estimate cannot be regarded as quantitatively reliable and, moreover, there are a number of factors favoring a raising of T_c that are not taken into account in the discussions given in[13,17], nevertheless, the restrictions on T_c that stem from the inequality $\varepsilon(0, \mathbf{k}) > 1$ seemed to be extremely serious and worrying.

From the results of this article it follows that such restrictions simply do not exist, because the inequality $\varepsilon(0, \mathbf{k}) > 1$ itself is incorrect for large k. We do not know any general requirements which would forbid an arbitrarily strong attraction between electrons at large values of k and, correspondingly, an arbitrarily high value of the critical temperature. Of course, the problem of the actual existence of structures with $\varepsilon(0, \mathbf{k}) < 0$ or of their artificial synthesis remains completely open, as yet. However, such a possibility cannot be excluded and it appears that the search for structures of this kind is an interesting and important problem of solid-state physics. It may be that the hopes of a radical raising of the critical temperature of the superconducting transition could be realized in precisely this way.

9. CONCLUSION

The outcome of the analysis performed in this article can be summarized in the form of the following statements:

a) The Kramers–Kronig relation for the quantity $1/\varepsilon(\omega, \mathbf{k})$ and the restrictions $\varepsilon(0, \mathbf{k}) > 1$ or $\varepsilon(0, \mathbf{k}) < 0$ are necessarily valid for any stable system;

b) the Kramers–Kronig relation for $\varepsilon(\omega, \mathbf{k})$ and the stronger restriction $\varepsilon(0, \mathbf{k}) > 1$ hold good only in the long-wavlength limit $k \to 0$;

c) for sufficiently large values of k, negative values of the static dielectric permittivity become permissible;

d) there exists in principle a class of substances within which the static interaction of the electrons has the character not of repulsion, as *in vacuo*, but of attraction.

Analogous conclusions as applied to the magnetic permeability of a substance are also valid.

On a more general level, the analysis carried out in this article serves as a caution against the mechanical extension of the formulas of ordinary electrodynamics to the case of media with spatial dispersion.

[9)] It is necessary to keep in mind the condition, not stipulated up to now, that the infinitesimal quantity δ appearing in many formulas of this article has a sign coinciding with the sign of ω. We also note that, for $\omega_0^2 < 0$, violation of the Kramers–Kronig relation for ε occurs because the function $\varepsilon(t)$ loses the property of causality: according to the first formula (18), in this case, $\varepsilon(t, \mathbf{k}) - \delta(t) \sim \exp(-\sqrt{\omega_0^2}\,|t|)$.

The author is grateful to his co-authors of the book[3] for numerous discussions, in particular to V. L. Ginzburg, on whose initiative this article was written, and also to L. V. Keldysh and A. D. Linde for a useful discussion of the problems covered.

[1]D. Pines and P. Nozières, The Theory of Quantum Liquids, Benjamin, N.Y., 1966 (Russ. transl. "Mir," M., 1967).
[2]P. C. Martin, Phys. Rev. 161, 143 (1967).
[3]L. N. Bulaevskiĭ, V. L. Ginzburg, G. F. Zharkov, D. A. Kirzhnits, Yu. V. Kopaev, E. G. Maksimov and D. I. Khomskiĭ, Problema vysokotemperaturnoĭ sverkhprovodimosti (The Problem of High-Temperature Superconductivity), Nauka, M., 1976.
[4]V. M. Agranovich and V. L. Ginzburg, Kristallooptika s uchetom prostranstvennoĭ dispersii v teorii eksitonov (Spatial Dispersion in Crystal Optics and the Theory of Excitons), Nauka, M., 1965 (English translation published by Interscience, N.Y., 1966).
[5]V. P. Silin and A. A. Rukhadze, Elektromagnitnye svoistva plazmy i plazmopodobnykh sred (Electromagnetic Properties of Plasmas and Plasma-like Media), Gosatomizdat, M. 1961.
[6]D. A. Kirzhnits and V. N. Sazonov, p. 84 in the book "Einshteĭnovskiĭ sbornik," ("Einstein Memorial Volume") Nauka, M., 1974; P. L. Chonka, *ibid.*, p. 178.
[7]D. A. Kirzhnits, V. Ya. Feinberg, and E. S. Fradkin, Zh. Eksp. Teor. Fiz. 38, 239 (1960) [Sov. Phys. JETP 11, 174 (1960)].
[8]L. D. Landau and E. M. Lifshitz, Elektrodinamika slposhnykh sred (Electrodynamics of Continuous Media), Gostekhizdat, M., 1957 (English translation published by Pergamon Press, Oxford, 1960).
[9]M. A. Leontovich, Vvedenie v termodinamiku (Introduction to Thermodynamics), Gostekhizdat, M.-L., 1952; Statisticheskaya fizika (Statistical Physics), Gostekhizdat, M.-L., 1944.
[10]L. D. Landau and E. M. Lifshitz, Statisticheskaya fizika (Statistical Physics), Nauka, M., 1964 (English translation published by Pergamon Press, Oxford, 1969).
[11]S. Coleman and E. Weinberg, Phys. Rev. D7, 1888 (1973); D. A. Kirzhnits and A. D. Linde, Zh. Eksp. Teor. Fiz. 67, 1263 (1974) [Sov. Phys. JETP 40, 628]; J. Iliopoulus, C. Itzykson and A. Martin, Rev. Mod. Phys. 47, 165 (1975).
[12]D. A. Kirzhnits, Yu. E. Lozovik and G. V. Shpatakovskaya, Usp. Fiz. Nauk 117, 3 (1975) [Sov. Phys. Uspekhi 18, 649 (1976)].
[13]M. L. Cohen and P. W. Anderson, p. 17 in Proc. AIP Conference on *d*- and *f*-band Superconductivity (Rochester, 1971), N.Y., 1972.
[14]D. A. Kirzhnits and Yu. A. Nepomnyashchiĭ, Zh. Eksp. Teor. Fiz. 59, 2203 (1970) [Sov. Phys. JETP 32, 1191 (1971)]; Yu. A. Nepomnyashchiĭ, Teor. Mat. Fiz. 8, 413 (1971) [Theor. Math. Phys. (USSR) 8, (1971)]; Yu. A. Nepomnyashchiĭ and A. A. Nepomnyashchiĭ, *ibid.*, 9, 137.
[15]Material from the International Conference on Quantum Crystals (Tbilisi, 1974), FNT (Low Temperature Physics) 5–7 (1975).
[16]E. G. Maksimov, Zh. Eksp. Teor. Fiz. 69, 2236 (1975) [Sov. Phys. JETP 42, 1138].
[17]D. A. Kirzhnits, E. G. Maksimov and D. I. Khomskiĭ, Preprint, IFM UNTs AN SSSR (Institute for Physics and Mathematics, Urals Scientific Center, Academy of Sciences of the USSR), 1973.
[18]N. N. Bogolyubov, V. V. Tolmachev and D. V. Shirkov, Novyĭ metod v teroii sverkhprovodimosti (A New Method in the Theory of Superconductivity), Izd. AN SSSR, M. 1958 (English translation published by Consultants Bureau, N.Y., 1959).

Translated by P. J. Shepherd.

A "paradox" of electrodynamics

B. M. Bolotovskiĭ and V. A. Ugarov

P. N. Lebedev Physics Institute of the Academy of Sciences of the USSR and V. I. Lenin Moscow State Pedagogic Institute

Usp. Fiz. Nauk **119**, 371–374 (June 1976)

PACS numbers: 03.50.–z

1. The application of Gauss's theorem to a nonstationary spherically symmetric charge distribution leads to an unexpected result. Suppose that the charge is bounded by a sphere of radius R, and the charge density inside the sphere is determined by the function $\rho(\mathbf{r}, t)$. If we seek the field outside the charged sphere, by Gauss's theorem we obtain

$$4\pi r^2 \varepsilon E(r, t) = 4\pi Q(t), \tag{1}$$

where $Q(t)$ is the total charge of the sphere at time t. From (1) it follows immediately that

$$E(r, t) = \frac{Q(t)}{r^2}, \tag{2}$$

and for the potential $\varphi(\mathbf{E} = -\nabla\varphi)$

$$\varphi(r, t) = -\frac{Q(t)}{r} \qquad (r > R). \tag{3}$$

The results (2) and (3) are, of course, surprising. Gauss's theorem is, in essence, the Maxwell equation $\operatorname{div} \mathbf{D} = 4\pi\rho$, and in the Maxwell theory the speed of propagation of interactions (the speed of propagation of the field) is finite and equal to the speed of light $1/\sqrt{\varepsilon\mu}$. From (2) and (3) it is possible to conclude that the field propagates instantaneously. Instantaneous propagation of the field manifestly contradicts the special theory of relativity.

The fact that the expression (3) corresponds to infinitely fast propagation of the interaction can also be seen, in particular, from the fact that D'Alembert's equation (for the vacuum)