538.56

QUASIOPTICAL METHODS OF GENERATION AND TRANSMISSION OF MILLIMETER WAVES

B. Z. KATSENELENBAUM

Usp. Fiz. Nauk 83, 81-105 (May, 1964)

1. INTRODUCTION

THE term quasioptics covers a group of problems which pertain to devices used for the production of electromagnetic wave fields, and in which, on the one hand, methods of geometrical optics are used (focusing, refraction in prisms and lenses) and on the other hand the decisive role is played by diffraction phenomena.

In ordinary optical instruments, diffraction is of secondary significance, since the main geometrical parameter of the instrument, the beam width, is much larger than the wavelength. In cases where diffraction is significant, geometrical methods can usually not be employed. On the other hand, at centimeter or longer wavelengths, the situation is reversed: the beam cross section and the wavelength are of the same order of magnitude, and geometrical-optics phenomena are of relatively little importance. Only the simplest properties of ordinary cavity resonators and waveguides can be described by using the asymptotic form of the wave equation (geometrical optics) alone.

In the centimeter region there is one essential exception, namely when large antennas are used, but in the present review, which is devoted to new applications of optical methods to waveguides, we shall in effect disregard such antennas (see, incidentally, Sec. 4, Item 5), which have been in use for a long time and are described in the literature in detail (see, for example, [1]).

The practical utilization of millimeter and submillimeter radio waves has raised at least three new technical problems, which are quasioptical in the sense indicated above. This article is essentially devoted to an examination of these problems. They can be used to illustrate the results of the overlapping of the formalisms of geometrical optics and wave theory, which is most interesting from the methodological point of view. The first of these problems deals with mirrors and analogous devices (prisms, lenses) in very broad waveguides of round cross section, in which only the H_{01} mode propagates, the second problem deals with the theory of open resonators, which are widely used in lasers, while the third deals with lens lines or their analogs, mirror lines, over which radio waves are transmitted with low loss (per unit length) by multiple relaying of a relatively narrow beam.

From the mathematical point of view quasioptics leads to some special problems in the asymptotic theory of diffraction. In the first of the three problems we must combine methods of optics and waveguide theory. A feature of the second and third problems is that the ratio of the transverse beam dimension to the wavelength, that is, the large parameter of the problem, turns out to be not very large, and there exists still another, geometrical parameter which is of the same order of magnitude. This large geometrical parameter is the ratio of the distance between the elements of the apparatus, that is, the beam length, to its width. A typical situation in these two problems is to determine the field in the Fresnel zone in the case of diffraction by a large body. Mathematically the second and third problems are almost identical. We shall not consider here many devices [2,3] such as a prism power splitting device or any interferometer, in which the optical methods have been transferred to the millimeter wave region without a modification in principle.

Some problems in quasioptics have been dealt with long ago in the theory of ultrasound, where similar relations obtain between the geometrical dimensions of the apparatus and the wavelength. However, radiophysics has apparently introduced many new ideas into this field.

2. OPTICAL ELEMENTS IN BROAD WAVEGUIDES

1. Symmetrical magnetic waves in a waveguide of round cross section have at high frequencies very low ohmic losses, which decrease monotonically with increasing frequency. This simple result of waveguide theory has been the basis of development of channels for the transmission of signals in a very broad frequency band, in a round waveguide operating in the H_{01} mode. The present status of work on this problem (see, for example, [5,6]) gives grounds for assuming that this is technically feasible; the possibility that this method can be used to transmit not only signals but also energy is likewise not excluded [7]. Such a channel will have low ohmic losses at large values of the parameter ka, where k is the wave number of the propagating wave and a is the radius of the waveguide. Reasonable values of ka, determined by different considerations, lie apparently in the region $ka \approx 30-50$ and above.

One of the main difficulties in this problem is the stabilization of the H_{01} mode in different irregular sections. Any irregularity will cause, generally speaking, part of the H_{01} mode energy to go over into other

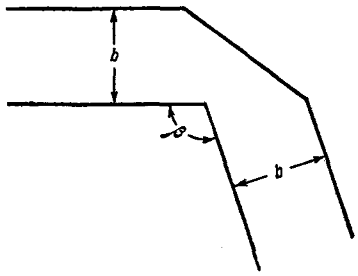

FIG. 1. Plane mirror in the bend of a waveguide.

modes; this energy is lost and moreover causes signal distortion. Therefore the conversion of energy into undesirable modes should be made as small as possible.

Optical devices are used to reduce such conversion losses, which arise whenever the waveguide axis changes direction. They have been proposed and even realized relatively long ago [8,9]. The simplest of these is a flat mirror installed in the waveguide bend in such a way that the normal to the mirror bisects the angle between the waveguide axes (Fig. 1, b = 2a). The use of such a mirror is based on a very simple idea: since ka $\gg 1$, the H_{01} field is close to the field of a plane wave locally, that is, in any region which is small compared with a. A plane wave incident on a flat mirror is reflected without distortion. Consequently the H_{01} wave should also be reflected with practically no distortion, that is, be deflected and proceed along the second waveguide. A similar geometrical-optics consideration enables us to expect a dielectric prism placed in the waveguide band (Fig. 2) to swing the front of the incident wave in such a way that the wave will pass almost completely into the second waveguide.

These considerations describe correctly the operation of such devices. The peculiarity of the problem, however, lies in the fact that the greatest interest indeed attaches to determining the corrections due to the finite nature of the parameter ka, and the associated diffraction effects. It is these corrections which describe the conversion losses in such devices.

At the present time we can regard as settled only the question of the conversion losses in a bend with a mirror (see Fig. 1). This question was considered experimentally in [8-9] and particularly in [10], while the theory is given in [11-13]; the agreement between theory and experiment is satisfactory.

2. The principles of the simplest theory [11] of a bend with a mirror are best developed using the auxiliary problem of a broad slot in the wall of a broad waveguide (Fig. 3). The presence of the slot gives rise to parasitic waves and to escape of some of the

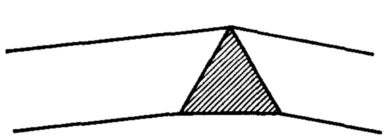

FIG. 2. Dielectric prism in the bend of a waveguide.

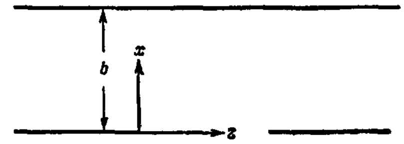

FIG. 3. Broad slot in the wall of a broad waveguide.

energy from the waveguide. The determination of the field in the right-hand waveguide, and, in particular, the determination of the amplitude of the fundamental wave (that is, the same wave which is incident from the left) are carried out in two stages. At first the field is determined on the slot, followed by determination from this field of the field in the right-hand waveguide.

The field on the surface of the slot is determined (in the highest order in the small parameter) in elementary fashion; knowledge of this field enables us to determine all the essential diffraction characteristics and, what is particularly important, the corresponding formulas result in small errors if the inaccuracy in the expression for the field on the slot is small. This does not occur when the field in the waveguide section beyond the slot is calculated directly from the field in the section ahead of the slot. In order to be able to determine the waveguide field from the field in the waveguide section it is necessary to know the field in the section very accurately, much more accurately than when the waveguide field is determined from the field on the continuation of its walls. that is, on the slot.

The field on the slot is best determined from a comparison with the Sommerfeld problem of diffraction on a half-plane. To this end it is necessary to separate the term corresponding to the Brillouin wave from the waveguide wave which is incident on the slot from the left. For example, for a plane waveguide of width b, with a boundary condition of the first kind, for an incident wave

$$u^{\text{inc}} = \sin \alpha_1 x \cdot e^{-ih_1 z}, \qquad \alpha_m = \frac{\pi m}{b},$$

$$h_m = \sqrt{k^2 - \alpha_m^2}, \quad u|_C = 0, \tag{2.1}$$

the field on the slot is equal to

$$u = \alpha_1 \sqrt{\frac{z}{2\pi k}}\, e^{-i\left(kz + \frac{\pi}{4}\right)}. \tag{2.2}$$

The z axis is directed here along the waveguide axis, the x axis is transverse to the waveguide, and the z coordinate in the amplitude is measured from the point of intersection of the z axis and the edge of the slot. For a boundary condition of the second kind, if

$$u^{\text{inc}} = e^{-ikz}, \qquad \frac{\partial u}{\partial n}\Big|_C = 0, \tag{2.3}$$

we have on the slot

$$\frac{\partial u}{\partial x} = \sqrt{\frac{k}{2\pi z}}\, e^{-i\left(kz + \frac{3\pi}{4}\right)}. \tag{2.4}$$

The field on the slot is determined in both cases by the incident-wave field not over the entire surface of the front, that is, not over the entire exit aperture of the left-hand waveguide, but only on a small region adjacent to the boundary of the slot. The dimension of this region is determined by the extent of several Fresnel zones for the most remote points of the slot. If the length of the slot is l, then the dimension of this region is equal to μb, where

$$\mu = \left(\frac{\pi l}{kb^2}\right)^{1/2} \tag{2.5}$$

is the principal small parameter of the problem. The condition

$$\mu \ll 1, \tag{2.6}$$

which we shall assume to be satisfied, signifies that the slot is exposed to only a small part of the front of the incident wave. This enables us to use the solution of the diffraction problem on a half-plane, and subsequently transfer the solution of the auxiliary problem (see Fig. 3) to the mirror problem (see Fig. 1).

In exactly the same way, if a symmetrical magnetic wave with a Hertz vector

$$J_0\left(\frac{\nu r}{a}\right)\exp(-ih_1 z), \qquad \nu = 3.83,$$

is incident in a waveguide of round cross section, we obtain on the slot, from the expression for the azimuthal component of the electric field of the incident wave (which we also denote by u),

$$u = \frac{\nu^2}{a^2\sqrt{2\pi}} J_0(\nu)\sqrt{kz}\, e^{-i\left(kz-\frac{\pi}{4}\right)}. \tag{2.7}$$

Knowing the field on the slot, we can determine the amplitude of the main and parasitic waves moving into the right and left waveguides from the usual formulas of the theory of excitation of a waveguide by specified magnetic and electric surface currents. At the same time, we determine from the same Sommerfeld problem the magnetic field on the slot, and then the energy flux into the slot. For an H_{01} mode, the energy loss of the fundamental wave is proportional to μ^3; the same takes place in the analogous two-dimensional scalar problem with boundary condition of the first kind (2.1). For a boundary condition of the second kind (2.3) and the analogous problem of the E_{01} mode in a round waveguide, the losses are much larger and are proportional to the first power of the small parameter, not the third. Half of the energy loss goes through the slot, and the other half is distributed among the parasitic waves that move forward, forming a broad wave-number spectrum. The losses for the problem (2.3) are larger than for problems (2.1) and (2.7) because in the case of a boundary condition of the second kind more energy is transported in the incident wave near the walls than in the case of the boundary condition of the first kind.

The foregoing method comprises a combination of the optical (determination of the field on the slot) and waveguide formalisms. It apparently corresponds to the physical nature of the problem, which is also of intermediate character: on the one hand there are very many waveguide waves and the determination of the amplitude of each undesirable wave is unnecessary and impossible, while on the other hand the structure of the transmitted wave should be as close as possible to the structure of one of the waveguide waves.

3. The auxiliary problem concerning a broad slot in a wall of a flat waveguide with boundary condition of the second kind (2.3) was solved in [12] by a more rigorous method, of the so-called successive diffractions.

Assume that at first there is only the left-side waveguide, and the right-side waveguide is removed. The field that results from this is known; it was determined [14,15] in papers on the theory of diffraction by the open end of a waveguide. We then place the right-side waveguide in this field and remove the left-side waveguide. In this case, too, we can determine the field by solving the problem of the diffraction of a compound wave (as found in the first step) by the open end. We then place in this field the left-side waveguide, assuming that there is no right-side waveguide, etc. This method actually imitates the process of formation of the field on the front of an incident pulse. It is known from the theory of diffraction by a slot in a plane screen, where the initial solution is the field produced by diffraction on a half plane; in [12] this initial solution was the field in the corresponding waveguide problems [14]. Several analytic expressions were obtained in [12] for the total field. In particular, several terms of the expansion of the energy of the transmitted fundamental wave in powers of μ were obtained. If the width of the waveguide is equal to the width of the slot, the losses of this wave are equal to

$$\frac{1}{\pi}\mu + \frac{1}{2\pi^2}\mu^2 + C\mu^3, \tag{2.8}$$

where C is made up of sine and cosine functions, whose arguments contain μ^2 in the denominator. The first term of (2.8) coincides with the result of the elementary theory [11]. The agreement between the exact theory and the approximate one remains satisfactory up to $\mu \sim 0.4-0.5$, but at larger values of μ the approximate theory gives incorrect results.

4. The flexibility of the method developed in Item 2 makes it easy to transfer it to the problem of a bend with a mirror, which, obviously, is equivalent to the problem of the waveguide cross, which results if the waveguides of Fig. 1 are reflected in the plane of the mirror. In this case two waves travel towards the crossing point. The wave traveling from the left, for example, experiences approximately the same disturbance as in a waveguide having slots that coincide with the mouths of the upper and lower waveguides. The energy which in the problem of Fig. 3 goes out of the slot and produces a space wave, creates in the cross para-

sitic waves in the lower waveguide. Several waves are formed—a narrow wave-number spectrum, which is shifted relative to the broad spectrum formed by the direct wave.

This yields the following result [11], which already pertains directly to the bend with the mirror. For a flat waveguide with wave (2.1), the losses are equal to $4\mu^3/3$; for a flat waveguide with wave (2.3), they are equal to $2\mu/\pi$, where the quantity l entering in μ (the width of the slot in the equivalent problem) is equal to $b/\sin\beta$. The H_{01} mode losses in a round waveguide are

$$\frac{8.7}{(ka \sin \beta)^{3/2}}. \quad (2.9)$$

In all cases this energy is uniformly distributed among the parasitic waves. One-quarter of the energy travels backward in the form of several high-number parasitic waves and an equal fraction goes downward in the form of similar waves; half of the energy goes into the lower waveguide in the form of many parasitic waves (broad number spectrum).

The higher the frequency, that is, the larger ka, the smaller the conversion losses which have, essentially, a diffraction character. The dependence on the frequency ($k^{-3/2}$ or $k^{-1/2}$) is quite complicated and unusual in the ordinary (non-waveguide) asymptotic problems of diffraction theory. It is not clear at present whether this dependence is general in any way and whether it holds, for example, for a bend with a prism (see Fig. 2).

The result in (2.9) shows that the construction of Fig. 1 is satisfactory even at not too high frequencies. For example, when ka ~ 60 approximately 2% of the incident energy is lost in a (right angle) bend with a mirror. The more the bend angle deviates from $\pi/2$, the broader the slot that the wave must overcome and the higher the conversion losses. For small angles, lower losses are probably obtained by using a bend with a prism.

5. The formalism used in [11] and [12] makes use of the asymptotic character of the problem from the very outset. This yields, of course, a solution which is valid only when $kb \gg 1$ (or, what is practically the same, $\mu \ll 1$); in addition, many features of the phenomena remain unclear. The problem of a right-angle bend in a rectangular waveguide with a mirror, under a boundary condition of the first kind, was investigated in [13] from the beginning to the end as a waveguide problem, using methods developed in the theory of planar waveguide junctions. The fields in the left-side and in the lower waveguides were expressed as sums over all the waves outgoing from the bend, and one wave (2.3) incident on the bend with unit amplitude. A system of an infinite number of linear algebraic equations was set up for the unknown amplitudes of the outgoing waves.

A detailed numerical analysis has shown that to obtain a sufficiently accurate numerical result it is necessary to take into account all the propagating waves as well as two or three nonpropagating waves. Of course, the higher the frequency, the more laborious this method is, and it apparently does not yield simple analytic estimates for a large value of the parameter kb. However, unlike the asymptotic methods, the results obtained in [13] are exact.

In practice the calculations could be carried through up to $kb = 10\pi$. Figure 4 shows the energy loss of the fundamental wave as a function of kb. The dashed line corresponds to the asymptotic solution, according to which the losses are equal to $4\pi^{3/2}(kb)^{-3/2}/3$. The asymptotic solution already gives satisfactory results for $kb > (4-5)\pi$.

Probably the most interesting result is the behavior of the amplitude of the fundamental wave near the frequencies ($kb = n\pi$) at which new parasitic waves arise. The corresponding resonance phenomena drop out of consideration completely in the asymptotic solutions; prior to publication of [13], the prevalent opinion was that the waveguide resonances are always weak. Actually it turned out that when the frequency of the parasitic waves goes through the critical value, no essential perturbation of the fundamental wave is produced (at any rate in the system under consideration).

For more complicated installations (for example, for a bend with a mirror in a round waveguide or for a prism, lens, etc.), so rigorous a waveguide approach is much more cumbersome, although it probably is still feasible. More promising, however, is the development of a special asymptotic formalism, which corresponds more to the intermediate quasioptical character of the problems. These asymptotic methods are of particular interest in conjunction with investigations such as [13], which ensure a continuous transition from the exact (numerical) to the approximate methods.

In conclusion we note that the simplicity of the physical idea on which different optical devices in waveguides are based has led and should continue to lead to the appearance of many clever technical suggestions (see, for example, [16]). One of the causes of the slow development of the technology in this field is the lack of a sufficiently general and at the same time effective and simple mathematical theory of quasioptical phenomena in broad waveguides.

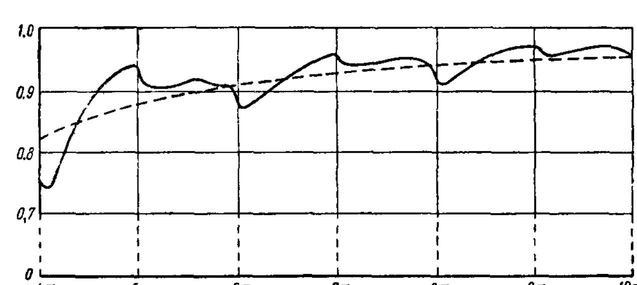

FIG. 4. Energy of wave transmitted through the bend of a plane waveguide with a mirror[13].

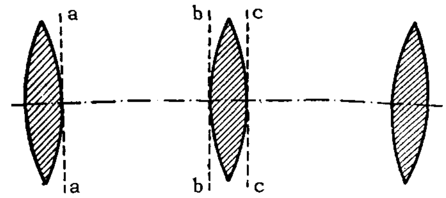

FIG. 5. Lens line.

3. BEAM WAVEGUIDE AND OPEN CAVITY

1. The suggestion that millimeter radio waves be transmitted along a line consisting of a series of successively placed dielectric lenses, the dimensions of which are large compared with the wavelength, was first formulated a few years ago [17,18]. Such a lens line (Fig. 5) is based on an elementary idea. The radio waves from the source strike the lens, pass through it, propagate to the next lens, etc. Every time the segment between lenses is traversed, some diffraction divergence or "spreading out" of the beam takes place, as a result of which the wave arriving at the lens has a structure which differs from the structure of the wave leaving the preceding lenses. The purpose of the lens is to compensate for this change in the structure and to restore the field distribution in the beam or, more accurately, the phase distribution over the section of the beam, since the distribution of the field amplitude remains practically unchanged.

In addition, part of the energy emerging from a given lens travels sideways and does not reach the next lens at all, and this leads to radiation energy losses. If we add to these the losses in the lens material and the reflection losses, as well as the losses connected with the fact that real systems are not ideal geometrically, we find that the line is subject to running losses (per unit length). For a complete energy analysis of the line it is also necessary to take into account the losses on the receiving and transmitting ends.

Let us consider the quantities that characterize the action of such a lens. We denote the distance between lenses by L, the cross section of the lens being characterized by a quantity a. For round lenses a is the radius of the lens, and for the two-dimensional case under consideration 2a is the width of the lens. It is assumed that $ka \gg 1$ and $L/a \gg 1$. The parameter on which the radiative losses depend primarily can be obtained from elementary considerations. The beam, the width of which is of the order a, has in the case of uniform distribution an angular diffraction divergence on the order of 1/ka. At a distance L this leads to the broadening of its transverse dimensions by L/ka; the relative broadening is L/ka^2. Unless measures are taken to minimize the energy outgoing from the peripheral part of the beam, each segment of length L would be accompanied by radiative losses of the order of L/ka^2. We shall show that it is possible to reduce these losses by suitably shaping the beam and making it nonuniform. To make the radiation losses small, the quantity

$$c = \frac{ka^2}{L} \quad (ka \gg 1, \ L \gg a) \tag{3.1}$$

must not be small. It is found that c can be only slightly smaller than 2π.

Thus, in order to make the radiation losses small, the lenses must be placed as close to one another as possible, the beam radius must be made large, and the wavelengths must be as short as possible. Each of these requirements introduces its own complications. If the lenses are very frequent, the running dielectric losses and the reflection losses are large. If the radius of the beam is large, the conditions for its propagation must be maintained within a very large volume. As to waves that are shorter than several millimeters, they are barely coming into extensive use, in view of problems in the development of transmission lines for such waves. In the submillimeter band there are also losses in air [19], in long lines it is necessary to provide an artificial medium or vacuum, etc. Several isolated reports were published on the development of long-distance lens-type transmission lines. In [20] the following parameters are proposed: wavelength 3.3 mm, a = 15 cm, L = 47 meters. The expected losses are of the order of 3 dB/km. The entire line must be contained in a concrete pipe. In [21] is described a line containing 40 lenses per kilometer, with a loss of 1.2 dB/km at 8 mm wavelength. An investigation of short lines is reported in [8].

2. The purpose of the lenses is to correct the cross-sectional phase distribution of the incoming wave. More accurately, if the field of the wave (any one of its components) arriving at the lens is $u(x, y)$, then the outgoing wave should have a field $u(x, y) \exp[i\varphi(x, y)]$. The function $\varphi(x, y)$ characterizes the phase correction, which is ensured by the difference between the optical lengths of the rays passing through different points of the lens. In other words, a diverging beam approaching the lens should be focused by the latter and changed again into a converging beam. The transverse diffusion occurring between the two lenses [22] again causes a diffraction divergence of the beam, which should be compensated for by the focusing action of the next lens. The quasioptical nature of the process is probably most clearly illustrated by the fact that it is necessary to use both diffraction and geometrical-optics concepts to describe it.

Converging lenses (dielectric or metal-dielectric) are, of course, not the only possible phase correcting means. Focusing reflecting mirrors of different types (Fig. 6a–b), proposed for this purpose in [23] and [24], can also be used. The rays incident on different points of the mirror pass through different paths, and this produces the required phase correction. An obvious advantage of such correctors, compared with lenses,

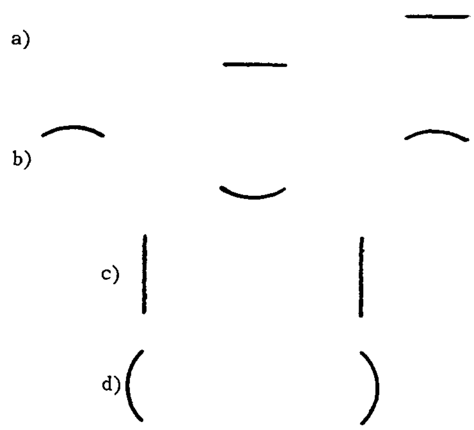

FIG. 6. a–b) mirror transmission lines; c) open resonator with plane mirror; d) open resonator with focusing mirrors.

is the absence of dielectric losses or analogous (and likewise considerable) losses in the artificial dielectric. The losses due to the conductivity of the mirrors can be made very small by using waves that are polarized parallel to the mirrors and strike the latter at a small glancing angle. If, in addition, the mirrors are made in the form of bodies of revolution, then the entire system, namely the sequence of barrel-like mirrors with a common axis [25], will have a field structure very similar to that of a round waveguide with a symmetrical magnetic wave. A shortcoming of such a system is the smallness (in the case of glancing incidence of the wave) of the effective beam width [the value of a in (3.1)]; small radiation losses can therefore be ensured only for submillimeter and shorter waves. In systems in which the angle of incidence is of the order of $\pi/4$ [24,16], the mirrors can be spaced far apart.

3. A physical analogy exists between lines made up of several phase correctors and the open resonators with plane (Fig. 6c) or curved (Fig. 6d) mirrors, which were proposed somewhat earlier [26-28]. The mathematical formalism which describes these two types of systems is in first approximation identical; after reflection from the right-hand mirror, the wave in the resonator moves to the left-hand mirror, and the wave moves in the line, after passing through each lens (or after being reflected from each mirror), to the next lens (or mirror). If we speak not of mirrors or of lenses, but of phase correctors, then the description of the processes in the lines and in the resonators becomes identical—the wave passes in succession between correctors, and experiences phase correction in each. The oscillation whose waveform is duplicated in each succeeding corrector, and whose amplitude decreases exponentially, corresponds in a resonator to the natural oscillation, and in the line to the natural wave. The decrease in the oscillation energy on going from one corrector to the next gives in transmission-line terms the running attenuation over a length L, and in resonator terms the damping per passage.

Resonators are widely used in lasers (optical quantum generators), where they serve as the oscillating system, and in measuring techniques, as microwave analogs of the optical Fabry-Perot interferometer. The growing interest in them is due primarily to their use in lasers.

From the technical point of view, the requirements imposed on lines and resonators are not the same. In long lines it is particularly important to ensure small running losses, while in short lines it is necessary to reduce the losses at the end of the path. A large beam radius is a shortcoming for the line, whereas for resonators the dimension of the mirror is not so important. In addition, the medium inside the resonator is optically inhomogeneous, and this limits the application of the relatively simple theory which will be described below.

For open resonators, a very important factor is the relative poverty of the spectrum [29] compared with ordinary (that is, closed) resonators. For lines this means large attenuation of almost all the higher natural waves—a property which perhaps makes it possible to employ the relatively simple frequency modulation for communication [20]. Fabry-Perot interferometers have their own specific problems (see, for example, the literature in [30]); as in the case of resonators, the quality of their mirrors is most important.

In presenting below the results of the mathematical investigation of the described systems, we shall be interested first of all in the radiation losses. This includes the most interesting problems, dealing with the relation between the geometrical-optics and diffraction theories. It is precisely here that the most important results have been obtained in recent years, although on the whole the problem cannot be regarded as solved even mathematically. Furthermore, in all the technical variants (Figs. 5 and 6) it is necessary to obtain small radiation losses, although for quantum generators this is not so important as for transmission lines, particularly long lines.

4. The most flexible and effective mathematical formalism for the investigation of the problem consists in the following. We introduce as the unknown function the field at the output of one of the correctors. From this we determine the field at the input of the next corrector. The field at the output of the latter is obtained by a multiplication by $\exp[i\varphi(x,y)]$. The requirement is then made that the field at the output of the second corrector differ from the field at the output of the first corrector only by a constant factor (independent of x and y). This leads to a homogeneous Fredholm integral equation for the sought field.

The field at $z = L$ is determined from the field at $z = 0$ by using the Huygens principle. Assume at first that we are dealing with a lens line. We denote by $u(x,y)$, for example, the x-component of the magnetic field at $z = +0$. The field outside the corrector is assumed equal to zero. For the same component of the magnetic field at z = L − 0 we get

$$v(\xi,\ \eta)|_{L=z-0} = A\int u(x,\ y)\frac{e^{-ikR}}{R}\,dx\,dy, \qquad |A| = \frac{k}{2\pi}. \quad (3.2)$$

In two-dimensional problems $|A| = \sqrt{k/2\pi}$, and R in the denominator must be replaced by $R^{1/2}$. The integration is carried out here over the aperture of the first corrector (line aa on Fig. 5), the field is obtained at the input of the second corrector (line bb on Fig. 5), and

$$R^2 = \sqrt{L^2 + (x-\xi)^2 + (y-\eta)^2}. \quad (3.3)$$

We further replace R in the denominator by L, and R in the numerator by

$$R = L + \frac{(x-\xi)^2}{2L} + \frac{(y-\eta)^2}{2L}. \quad (3.4)$$

This is valid if

$$L^3 \gg ka^4 \quad (3.5)$$

a condition satisfied when $ka \gg 1$ and $c/2\pi \sim 1$. We substitute (3.4) in (3.2) and multiply by $\exp[i\varphi(\xi,\eta)]$, and then obtain the field at the output of the second corrector, that is, at z = L + 0 (line cc on Fig. 5). If we stipulate then that

$$v(\xi,\ \eta)|_{z=L+0} = \lambda u(\xi,\ \eta)|_{z=+0}, \quad (3.6)$$

we obtain an integral equation for the function u:

$$e^{i\varphi(\xi,\ \eta)}\int u(x,\ y)K(x,\ y,\ \xi,\ \eta)\,dx\,dy = \lambda u(\xi,\ \eta), \quad (3.7)$$

where the kernel, in accordance with the foregoing, is

$$K = A\frac{e^{-ikL}}{L}e^{-i\frac{k}{2L}[(x-\xi)^2+(y-\eta)^2]}. \quad (3.8)$$

The quantity λ has a simple meaning; according to (3.6) it is equal to the decrease of the field on going from lens to lens. The radiation losses are equal to $1 - |\lambda|^2$. Thus, the losses of the different modes are determined by the moduli of the eigenvalues of (3.7), while the structure of the field is determined by the corresponding eigenfunctions, and the phase velocity is determined by the argument of the eigenvalues.

In place of (3.7), it is frequently convenient to consider for the field at the output of the lens the equation for the field in the central plane of the lens; we denote this field by w(x, y). From the definition of the function φ(x, y) it follows that

$$u(x,\ y) = w(x,\ y)\,e^{\frac{i}{2}\varphi(x,\ y)}. \quad (3.9)$$

Substituting this in (3.7) we obtain an integral equation

$$\int w(x,\ y)K(x,\ y,\ \xi,\ \eta)\,e^{\frac{i}{2}[\varphi(x,\ y)+\varphi(\xi,\ \eta)]}dx\,dy = \lambda w(x,\ y). \quad (3.10)$$

This method of analysis was initially proposed not for lens lines but for open resonators [27,28]. In the early papers [17,31] containing the theory of lens lines (called there beam waveguides), a different method was developed. Mathematically it leads to the same results, although via a more artificial and more complicated path. However, it emphasizes somewhat different aspects of the entire process. The sought field of the natural wave is represented in the form of an aggregate (beam) of plane waves, the direction of propagation of which lies inside a narrow cone. The unknown quantity is a function that gives the distribution of the amplitudes of these plane waves as a function of the angle between their propagation direction and the system axis (the z axis).

The variation of the wave field from z = +0 to z = L − 0 is described in this case simply by the laws of plane-wave propagation. The field at z = L + 0 is then obtained from the field at z = L − 0 by means of an operation which reduces essentially to multiplication by $\exp[i\varphi(x,y)]$ and a condition equivalent to (3.6) is stipulated.

A similar "beam method" was used also for open resonators in one of the earliest investigations [32], in which the two-dimensional problem of a resonator consisting of two flat mirrors was considered.

We present an analysis for both lines and resonators by the unified method described above, which leads to the integral equations (3.7) and (3.10).

We shall show that (3.10) describes also oscillations in an open resonator made up of two mirrors (Fig. 7), provided w(x, y) is taken to mean the magnetic field (or current) on the mirrors. Assume that w(x, y) is specified on the first mirror, and that w(ξ, η) on the second mirror is determined from the Huygens principle. Since the field on the first mirror is equal to zero, formula (3.2) will have for the field incident on the second mirror an additional factor ½. However, this factor disappears from the expression for the total field on the second mirror, since this mirror is also metallic and the magnetic field on it is double the incident-wave field. It is assumed in this argument that

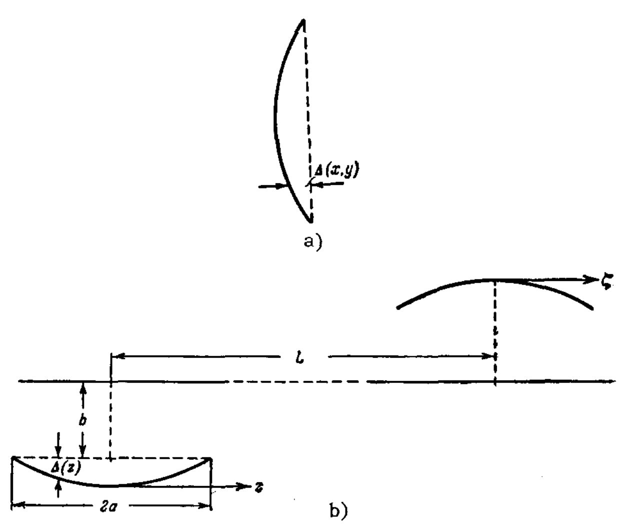

FIG. 7. a) Phase correction of focusing mirror in resonator; b) phase correction of focusing mirror in line.

the surfaces of the mirrors are nearly plane, so that the normals to the mirrors make everywhere small angles with the axis—a condition which we shall assume satisfied. The quantity R in (3.2) is now the distance between two points on bent mirrors. It differs from (3.4), which gives the distance between two points on the apertures, by terms equal to the distances between aperture and mirror points lying on one ray traveling from the opposite mirror (Fig. 7a). These distances can be replaced by the distances $\Delta(x, y)$ and $\Delta(\xi, \eta)$ between the aperture and mirror points having the same coordinates x, y or ξ, η.

A simple connection exists between Δ and φ:

$$\frac{1}{2}\varphi(x, y) = C - k\Delta(x, y), \quad \frac{1}{2}\varphi(\xi, \eta) = C - k\Delta(\xi, \eta), \tag{3.11}$$

where the value of the constant C is immaterial. The phase correction which occurs upon reflection from the bent mirror is due precisely to the fact that the rays reflected from different points of the mirror cover different distances $2\Delta(x, y)$. They acquire thereby additional phase shifts $2\Delta(x, y)k$; (3.11) then follows from the definition of the function $\varphi(x, y)$.

Thus, allowing for the additional terms (3.11) in the expression for the distance, we get for the current on the second mirror

$$\int w(x, y) K(x, y, \xi, \eta) e^{\frac{i}{2}[\varphi(x, y) + \varphi(\xi, \eta)]} dx\, dy, \tag{3.12}$$

where K is given in (3.8). The condition that the field be repeated on each mirror leads to (3.10).

If the function $u(x, y)$ is defined by the condition (3.9), then it represents the magnetic field on the plane aperture of the mirror. Thus, we obtain for this field u the same equations (3.7) as for the lens line.

Let us show, finally, that wave propagation in a mirror line is described by the same integral equations (3.7) or (3.10). We consider for simplicity a two-dimensional problem (the three-dimensional case is analyzed in [25]). Let the mirrors consist of bent ribbons, the sections of which by the plane y = const are shown in Fig. 7b; the fields do not depend on y. In determining the magnetic field $w(\zeta)$ on the second mirror from the field $w(z)$ on the first mirror by means of the Huygens principle, we should assume for the distance R in the exponential of (3.2) the expression

$$R = R_0 + \frac{L}{R_0}(\zeta - z) + \frac{2b^2}{R_0^3}(\zeta - z)^2 + \frac{2b}{R_0}[\Delta(z) + \Delta(\zeta)], \tag{3.13}$$

where $R_0^2 = 4b^2 + L^2$, and $\Delta(z)$ is equal to the distance between the mirror and the aperture point having the same value of z.

The difference between the theory of resonators (Fig. 7a) and the mirror line (Fig. 7b) is connected only with the difference in the expressions for R. The expression (3.13) contains, in comparison with (3.4), additional terms proportional to $\Delta(z)$ and $\Delta(\zeta)$. This means that each reflection from the bent mirror is accompanied by a phase correction; this phase correction is equal to, according to (3.13),

$$\varphi(z) = C - \frac{2b}{R_0}\cdot 2k\Delta(z). \tag{3.14}$$

This formula can, of course, be obtained from the usual geometrical calculation of the difference in the path between the two rays for inclined incidence on a bent surface. Formula (3.14) has the same meaning as (3.11); it enables us to determine the required shapes of the mirrors from the selected function φ.

The presence in (3.13) of a term linear in z and in ζ signifies that $w(z)$ contains a rapidly varying factor corresponding to wave propagation in the z direction. Let us separate this factor, that is, let us introduce the function

$$\overline{w}(z) = w(z)\, e^{i\frac{L}{R_0}kz}. \tag{3.15}$$

The integral equation for the slowly-varying function $\overline{w}(z)$ turns out to be precisely the same as for the current $w(x)$ on the mirrors of some resonator (two-dimensional problem), if the width of these mirrors is equal to the width of the beam, that is, $2a(2b/R_0)$. The latter statement follows directly from a comparison of the quadratic terms in (3.13) and (3.4). Introducing in the expression for the magnetic field on the aperture of the mirrors in place of $\overline{w}(z)$ the quantity $\overline{u}(z)$ in accordance with (3.9), that is, the factor preceding $\exp(ikLz/R_0)$, we obtain for this quantity the two-dimensional variant of (3.7).

Thus, (3.7) and (3.10) describe the fields in the lens and mirror lines and in open resonators. Their solutions—the forms of the eigenfunctions and the eigenvalues—are determined by the geometrical parameters and by the form of the function φ, that is, by the character of the phase correction. It is noted in [33] that, in general, problems involving the theory of equations such as (3.10), which have a complex symmetrical kernel, have been little studied.

5. The most interesting phase correction, which has been investigated in greatest detail, will be considered in the next section. We present here some results pertaining to the special case when

$$\varphi(z) \equiv 0, \tag{3.16}$$

that is, when no correction is made at all; in resonators and in mirror lines this takes place upon reflection from plane mirrors. The effect of each mirror on the structure of the wave consists here only of removing from the wave that part which goes past the corrector. A "lens" variant of such a system is possible and has even been experimentally realized [34]. It consists of diaphragms arranged one after the other in a row.

Resonators consisting of two flat mirrors were the first forms of open resonators used in lasers [26]. It could be assumed from the outset that the Q of such resonators would be very small. Indeed, as shown in

Sec. 1, when the current on the mirrors has uniform distribution, each reflection is accompanied by a large energy loss, of order $1/c$. It became clear, however, that the diffraction effects produced appreciable changes in the field structure, and a current distribution on the mirrors such as would ensure much lower losses. Finite mirrors separate from the aggregate, as it were, the plane waves which form the beam; the waves that travel at large angles to the axis rapidly leave the resonator, and all the remaining ones are reradiated many times.

Equation (3.10) for ribbon-type round (disc) mirrors was integrated [27] by successive iterations with an electronic computer. This imitated, essentially, the process of establishment of the oscillations; it was found that some 300 successive reflections were needed. The fields of the natural modes decrease quite rapidly towards the periphery of the mirror. For example, for the fundamental mode, at $c \approx 4\pi$ the field on the edge of the ribbon or disc is approximately one-quarter that at the center.

The radiation losses are smaller than $1/c$, but quite large. For example, for a ribbon operating at the fundamental mode the loss is 0.08 at $c = 2\pi$, 0.03 at $c = 4\pi$, etc. For a round disc the corresponding values are 0.18 and 0.07.

For rectangular mirrors, the distribution of the current as a function of each coordinate is independent and is determined by its own parameter c, in which a is half the length of the corresponding side. The losses connected with the finite size of the mirrors in each of the two directions are additive.

In [29] there was used an entirely different and at first glance paradoxical treatment of the entire process of formation of a field in an open resonator with plane mirrors. The resonator was regarded as a segment of a waveguide open on both ends (with a wave propagation direction shown vertical in Fig. 6c). The following result is known from waveguide theory: if the oscillation frequency is very close to the cutoff wavelength of the wave incident on the waveguide from its open end, then this wave will be reflected almost completely from the open end. Such a wave will be successively reflected in the waveguide segment from the upper and lower ends, forming standing waves. For a very broad waveguide (waveguide with L much larger than the wavelength), a mode with arbitrarily large wave number can occur. This will be possible not at all frequencies, but at some discrete frequencies close to the critical wavelengths in the waveguide of width L. The natural frequencies of the open resonator are thus separated. The most interesting in this case is the possibility of relating the radiation losses in the open resonator with the radiation coefficient from the open end of a waveguide, and determining the analytical form of the current on the mirrors. A comparison of the results of such a calculation with the exact values [27] obtained with an electronic computer from (3.10) has shown that they are very close. This formalism makes it possible to carry out the calculations also for a resonator with two plane circular mirrors.

4. CONFOCAL PHASE CORRECTORS

1. The plane mirrors considered in the preceding section, for which $\varphi \equiv 0$, are a degenerate form of a corrector. Waves whose field decreases in a direction perpendicular to z are produced in such mirrors only because the correctors are finite in size, that is, owing to diffraction effects. When $\varphi \equiv 0$ (3.10) does not have in an infinite region eigenfunctions that decrease at infinity.

An investigation of truly focusing correctors, which concentrate the beam near the axis because of the curvature of the mirrors or the lenses, can begin with an examination of infinite correctors. The larger the field concentration effected by the infinite correctors, the smaller the dimensions of the finite correctors of the same form need be to ensure propagation of the wave with low radiation losses. In other words, the finite-dimension corrector forms which result in the lowest losses should be the same as those of infinite correctors that ensure maximum concentration. This statement is not fully convincing at first glance, since it does not take into account the field redistribution caused by diffraction. We shall show later, however (Item 4), that under definite conditions this can be rigorously justified, so that this redistribution, which plays a decisive role in the absence of correctors (3.16), is not particularly significant if the correction is properly chosen.

Infinite correctors are described by Eq. (3.10) with infinite limits. Let us write it down for the two-dimensional problem in the form

$$A\,\frac{e^{-ikL}}{\sqrt{L}}\int\limits_{-\infty}^{+\infty} w(x)\, e^{i\frac{kx\xi}{L}}\, e^{\frac{i}{2}\left[-\frac{kx^2}{L}+\varphi(x)\right]+\frac{i}{2}\left[-\frac{k\xi^2}{L}+\varphi(\xi)\right]}\,dx=\lambda w(\xi). \tag{4.1}$$

We confine ourselves further only to the most thoroughly studied quadratic correctors, that is, we put

$$\varphi(x)=\nu\,\frac{kx^2}{L}. \tag{4.2}$$

The first eigenfunction of (4.1) has under condition (4.2) the form

$$w(x)=e^{-\alpha\frac{kx^2}{L}},\qquad \alpha=\frac{\sqrt{2\nu-\nu^2}}{2}. \tag{4.3}$$

This function will decrease as $|x|\to\infty$ only if ν lies in the interval $0<\nu<2$; this is correct also for the remaining eigenfunctions of (4.1).

The greatest concentration is attained if $\nu = 1$, that is, when the phase correction is

$$\varphi(x)=\frac{k}{L}\,x^2. \tag{4.4}$$

This phase correction compensates for the terms with the coordinates squared in the expression for the distance between two points of different mirrors. A similar elementary analysis shows that the best phase correction in three-dimensional problems (that is, for rectangular or round mirrors and lenses) is given by the function

$$\varphi(x, y)=\frac{k}{L}(x^2+y^2). \tag{4.5}$$

We can arrive at formulas (4.4) and (4.5) also in a different way, on the basis of intuitive geometrical-optical representations. We shall speak, for concreteness, of resonators. We stipulate that all the rays which emerge from some point on the left-side mirror be reflected in the right-side mirror and again be gathered at one point of the left-side mirror. In the ray treatment this means minimum energy dissipation and maximum energy concentration. This condition will be satisfied if the path length for all the rays emerging from a given point of the left-side mirror and arriving at another point of the left-side mirror is the same, that is, it is independent of the point on the right-side mirror from which the given ray was reflected. This condition also causes the phase correction to compensate the x^2 terms, etc., in R, so that this distance contains only the product of the coordinates

$$\frac{k}{L}(x\xi+y\eta). \tag{4.6}$$

From this we again obtain formulas (4.4) and (4.5).

Bent mirrors with phase correction (4.5) were also introduced in optical Fabry-Perot interferometers[35,36], although the intent here was not so much to decrease the radiation losses as to facilitate the adjustment of the instrument.

According to (3.11), the mirrors of the resonators, for which condition (4.4) or (4.5) is satisfied, represent, within the limits of accuracy of the entire analysis, parts (round or rectangular) of spherical surfaces with centers in the middle of the opposite mirror. Since the distance between the focus of the spherical mirror and its top is equal to half the radius, the foci of both mirrors coincide. Such a system is called confocal. In the two-dimensional case, that is, for mirrors in the form of ribbons, the mirrors should be parts of cylindrical surfaces whose axes lie in the middle of the opposite mirror. Formulas (4.4) and (3.14) make it easy to find also the shapes of the mirrors used in a mirror line.

In a lens line, the shapes of the lenses that ensure phase correction (4.5) depend on the dielectric constant ϵ of the material. The lens half-thickness d(r) depends on r in accordance with the formula[17]

$$d(r)=\frac{a^2-r^2}{2L(\sqrt{\varepsilon}-1)}, \qquad r^2=x^2+y^2, \tag{4.7}$$

which can be readily obtained by determining the optical path of the rays passing through the lens. Within the limits of the accuracy of the entire analysis, it is necessary here to carry out calculation for rays parallel to the axis of the system; this is the same approximation made in (3.11).

Such lenses are also confocal—the foci of two neighboring lenses lie in a single point located half-way between them. All rays from any one point on say, the first lens are gathered at one point on the third lens, and so on. In particular, the center of the first lens is focused by the second lens on the center of the third lens.

2. In this item we present the main results of an analysis of confocal resonators. The starting point, in accordance with the foregoing, is the integral equation (3.10) with phase correction (4.4) or (4.5).

We begin with two-dimensional problems. The integral equation, in which we introduce new variables

$$x'=x\sqrt{k/L}, \qquad \xi'=\xi\sqrt{k/L} \tag{4.8}$$

reduces to the form

$$\int_{-\sqrt{c}}^{\sqrt{c}} w(x')\, e^{ix'\xi'}\, dx'=\sqrt{2\pi}\chi w(\xi'). \tag{4.9}$$

Here c is defined in (3.1) and the eigenvalue χ differs from λ in (3.10) by a phase factor; the loss per passage is equal to $1-|\chi|^2$. When $c=\infty$, that is, for unbounded correctors, this equation has eigenfunctions

$$w_m(x')=H_m(x')\, e^{-\frac{1}{2}x'^2}, \tag{4.10}$$

where $H_m(x')$ are Hermite polynomials ($H_0=1$, $H_1(x')=2x'$, etc.), and the eigenvalues are $\chi_m=i^m$ ([37], No. 7.376.1). Equation (4.9) can be solved by expanding the kernel in the unknown function in terms of the functions (4.10). Then χ is determined from the condition of the existence of a non-trivial solution of the infinite system of linear equations for the expansion coefficients of the sought eigenfunction in the functions w_m. However, for Eq. (4.9) we know also[38] an explicit expression for the eigenfunctions and eigenvalues for any value of the parameter c. The eigenfunctions are proportional to the azimuthal wave functions of a prolate spheroid of parameter c and argument x/a,

$$w_m=S_{0m}\left(c, \frac{x}{a}\right), \tag{4.11}$$

and the eigenvalues are proportional to the radial functions of the argument 1:

$$\chi_m=\sqrt{\frac{2c}{\pi}}\, i^m R_{0m}^{(1)}(c, 1), \tag{4.12}$$

where S and R are defined in [38].

Figure 8 gives the loss per passage as a function of c for m = 0, 1, 2.

These results also pertain directly to rectangular mirrors. The losses due to the finite x and y dimensions of the mirrors add up, while the eigenfunctions are the products of the corresponding one-dimensional functions (4.11), each of which depends on the corresponding value of c.

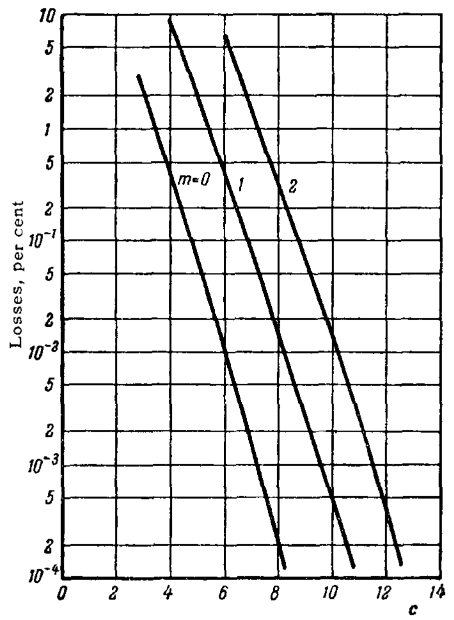

FIG. 8. Radiation losses for ribbon-type confocal corrector[38].

The field on the mirrors is close to the field on the infinite mirrors, and for small m it is described essentially by the exponential factor in (4.10):

$$w(x) \sim e^{-x^2/2x_s^2}, \quad \text{where } x_s = \sqrt{\frac{L}{k}}, \tag{4.13}$$

especially if a is noticeably larger than the width of the "light spot" x_s and the diffraction effects are small.

Inside the resonator the field decreases with increasing distance from the plane x = 0, approximately in accordance with the same law (4.13) as the current on the mirrors, but somewhat more rapidly. The effective size of the field is smallest in the central plane between the mirrors, where it is $\sqrt{2}$ times smaller than on the mirrors.

Expression (4.13) for x_s can also be readily obtained from elementary considerations. The dimension of the focal spot, as is well known, is equal to the wavelength divided by the sine of the angle at which the illuminated area is seen from the focus. If this area has a dimension x_s, then the size of the focal spot is L/kx_s. In the devices with a repeating field, which we are considering, the light spot should satisfy the same condition, that is, we should have $L/kx_s \approx x_s$.

The value of $\sqrt{c}$ is equal to the ratio of a to $(L/k)^{1/2}$. For large c in oscillations of the type (4.13), only a small part of the corrector will be illuminated. Of course, other oscillations are also possible, for which the entire or almost the entire corrector is illuminated. In terms of the theory developed here, this corresponds to oscillations of higher numbers; the light spot (that is, the region in which w(x) has the same order as w(0)) increases with the number m, in accordance with (4.10). Analogous conditions occur also in closed waveguides; when $ka \gg 1$, for example, a plane wave, that is, the flux of optical rays, falling on the mouth of a waveguide, generates an aggregate of higher-mode waveguide waves. When $c/2\pi \gg 1$, not only the fundamental oscillation (n = 0) but many of the higher modes have small radiation losses, and the structure of the propagating wave is determined to a great degree by the character of the excitation.

For confocal resonators with round mirrors, the phase correction is given by the function (4.5). The integral equation for the function $w(r, \vartheta)$ (ϑ—angle in cylindrical coordinate system) has solutions with arbitrary dependence on ϑ. By specifying this dependence we can carry out iteration with respect to ϑ in (4.10), which leaves a one-dimensional equation; for example, for the symmetrical oscillation, which, as turns out, has the smallest damping, w = w(r) satisfies the equation[27]

$$\lambda w\left(\frac{r_1}{r_s}\right) = \int_0^{\sqrt{c}} w\left(\frac{r}{r_s}\right) J_0\left(\frac{r}{r_s}\frac{r_1}{r_s}\right)\frac{r}{r_s}\, d\left(\frac{r}{r_s}\right). \tag{4.14}$$

The eigenfunctions of this equation[37] with $c = \infty$ are obtained by multiplying the Laguerre polynomials $L_n(r^2/r_s^2)$ ($L_0 = 1$, $L_1(x) = 1 - x$, etc.) by the function

$$e^{-r^2/2r_s^2}, \qquad r_s = \sqrt{L/k}. \tag{4.15}$$

The quantity r_s can be regarded as the radius of the light spot on the mirrors. Equation (4.14) and analogous equations for w(r) in the expression $w(r, \vartheta) = w(r)\exp(in\vartheta)$ were solved in [17,39] by expanding the unknown function and the kernel in series of the eigenfunctions of the equation for $c = \infty$ and by reduction to an infinite system of algebraic equations. In [27] this equation was solved directly by iteration. The results of both numerical methods coincide.

Figure 9 gives the percentage loss per passage for the fundamental wave as a function of c, while Fig. 10 gives the amplitude of the current on the mirrors as a function of r/a for several values of c[27]. These two figures (along with more detailed and more complete data pertaining to other waves, given in [27,38]), together with the formula (4.5), are essentially the central result of the theory of open resonators. They show, in particular, that when $c = 2\pi$ the radiation losses are quite small, but they increase very rapidly with decreasing c, approximately 100-fold when c is halved.

3. An analysis of unbounded phase correctors is useful also in investigations of finite non-confocal correctors with quadratic correction (4.2); if $a \geq x_s$, then the field on the finite and infinite correctors are nearly the same. For unbounded symmetrical quadratic correctors, the three-dimensional problem always reduces to a two-dimensional problem, to which we confine ourselves.

For two identical mirrors with radius of curvature ρ, the field is written out in (4.3), where we must put $\nu = L/\rho$. When $\nu = 1$, that is, for confocal resonators,

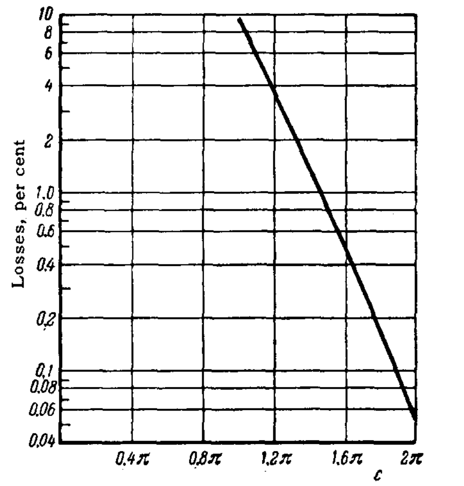

FIG. 9. Radiation losses for round confocal corrector[27].

(4.3) coincides, of course, with (4.10) with $n = 0$, or with (4.13). For $\nu = 0$ we get $\varphi(x) \equiv 0$ and we have the degenerate case of plane mirrors ($\rho = \infty$). Precisely the same conditions obtain also if $\nu = 2$, when the curvature of the mirror is twice as large as that of the confocal mirrors, that is, when their centers of curvature coincide. When $\nu < 0$ or $\nu > 2$, that is, for mirrors with even larger curvature or for convex mirrors, the field cannot be concentrated by infinite mirrors and no oscillations are produced.

The expression $2\nu - \nu^2$ is invariant under the substitution $\nu' = 2 - \nu$. Therefore the resonators corresponding to such values of ν and ν' have, in accordance with (4.3), the same field structure on the mirrors. The curvature of the mirrors of one of these resonators is larger than that of the confocal resonators, while that of the other is smaller. The width of the light spot on the mirrors of both resonators is the same and according to (4.3) it is equal to

$$\frac{\sqrt{L/k}}{\sqrt{2\nu-\nu^2}}. \tag{4.16}$$

Apparently we can assume, provided only a is larger than (4.16), that the radiation losses of such non-confocal resonators are equal to the losses of confocal resonators with the same ratio of a to the width of the light spot. In other words, the losses of non-confocal resonators can be determined from the Fig. 8, if we take c to mean the parameter $ka^2(2\nu - \nu^2)^{1/2}/L$, which is equal to

$$c=\frac{ka^2}{L}\left(\frac{2L}{\varrho}-\frac{L^2}{\varrho^2}\right)^{1/2}. \tag{4.17}$$

This assumption[38,40], which is not completely rigorous (it does not take into account the differences in the perturbation of the field by diffraction effects in the compared resonators) was verified in [41] for one value of c ($c = \pi$) by direct calculation. The dashed curve of Fig. 11 gives the loss as a function of $\nu - 1$, calculated in accordance with (4.17) and Fig. 8; the continuous curve was obtained by numerically integrating (3.10) subject to the condition (4.2).

A similar analysis can be made also for a resonator with two unequal mirrors. Assume that the first mirror has a radius of curvature ν_1 and the other ν_2, so that the phase corrections are determined by (4.2), in which $\nu_1 = L/\rho_1$ for the first mirror and $\nu_2 = L/\rho_2$ for the second. The functions $w_1(x)$ and $w_2(\xi)$, which give the currents on the first and second mirrors, are connected by the transformation

$$\lambda_1 w_2(\xi) = e^{\frac{i}{2}\varphi_2(\xi)} \int w_1(x)\, e^{\frac{i}{2}\varphi_1(x)} K(x, \xi)\, dx \tag{4.18}$$

and by a second such transformation from $w_2(\xi)$ to $w_1(x)$. From this we can obtain an integral equation for $w_1(x)$ or $w_2(\xi)$ [41]. It is simpler, however, to put immediately

$$w_1(x) \sim e^{-\alpha_1 \frac{kx^2}{L}}, \quad w_2(\xi) \sim e^{-\alpha_2 \frac{k\xi^2}{L}}. \tag{4.19}$$

We then obtain from (4.18) the simple expressions for

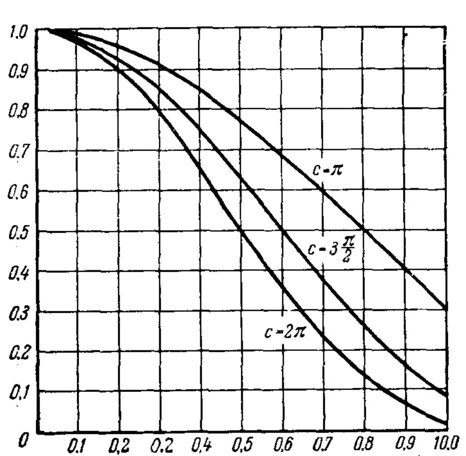

FIG. 10. Current on mirrors of confocal round correctors[27].

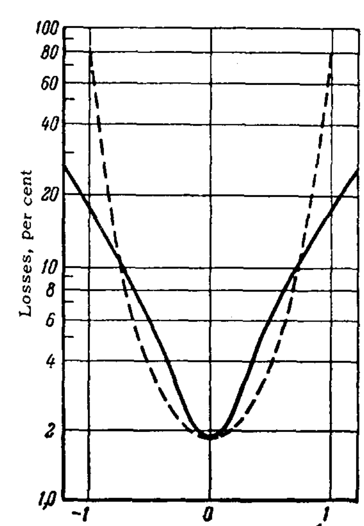

FIG. 11. Radiation losses for ribbon correctors[41].

α_1 and α_2

$$\alpha_1=\frac{\nu_1-1}{2}\sqrt{\frac{1}{(\nu_1-1)(\nu_2-1)}-1},$$

$$\alpha_2=\frac{\nu_2-1}{2}\sqrt{\frac{1}{(\nu_1-1)(\nu_2-1)}-1}. \qquad (4.20)$$

When $\nu_1 = \nu_2$, expressions (4.19) and (4.20) go over into (4.3). ν_1 and ν_2 will not be real for all values of α_1 and α_2. If the radicand in (4.20) is negative, then α_1 and α_2 are pure imaginary and the fields (4.19) in the infinite correctors do not decrease at infinity. In resonators with finite correctors, this corresponds to very large radiation losses. The oscillations have approximately the same character as in resonators made up of plane or concentric mirrors.

The region of values of the parameters ν_1 and ν_2, in which the infinite correctors form a beam of finite width, is determined in accordance with (4.20) by the conditions

$$0<(\nu_1-1)(\nu_2-1)<1. \qquad (4.21)$$

In the region (4.21), the obvious energy requirement $|\lambda_1\lambda_2| = 1$ is also satisfied. Condition (4.21) was first derived in [40] by a different method.

Figure 12 from [41] is drawn in coordinates $\nu_1 - 1$ and $\nu_2 - 1$. The shaded region is the one in which (4.21) is violated. At the points lying in this region, no oscillations are produced in the case of infinite correctors, while in infinite correctors the losses are appreciable. The curves on this figure show the attenuation for finite two-dimensional correctors ($c = \pi$); in the shaded area it increases much more rapidly on moving away from the confocal system.

An unexpected result is the instability of the confocal system. If the radii of curvature of both mirrors deviate from the condition $\rho = L$ in opposite directions ($\rho_1 > L$ and $\rho_2 < L$), then a nonfocusing system is produced (α_1 and α_2 are pure imaginary), and the limitation of the field in finite systems is due only to diffraction on the edges, which is accompanied by appreciable radiation losses. This effect was verified experimentally[42]. Physically, of course, no jump-like

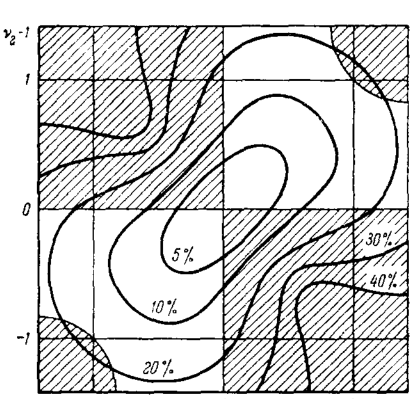

FIG. 12. Losses in resonator made up of different ribbon correctors[41].

changes in the oscillation conditions occur when a small change takes place in the curvature; an infinite corrector is a good idealization only in the case when the field in it decreases rapidly as $|x| \to \infty$. However, it is apparently advantageous to use resonators with somewhat larger or smaller mirror curvature than for confocal resonators, in order to guarantee against an accidental transition into the region of large losses.[40]

We note that the analysis of resonators with nonidentical mirrors does not enable us to draw any conclusions with respect to the line tolerances. Deviation of the form of one of the mirrors in the resonator from the theoretically calculated value is equivalent to identical deviation of the forms of all the even (or all the odd) correctors in the line, that is, to a case not realizable in practice. The question of tolerances in the lines should be solved independently; in particular, the problems of lines and resonators are no longer identical in this case.

4. The confocal correctors (4.4) and (4.5) have the following property: the eigenfunctions w of the integral equation (3.10) are real. This means in-phase mirror current (or field on the central plane of the lens). This circumstance emphasizes the geometrical-optical meaning of the assumed focusing. All the rays emerging from the different points, for example the left-side mirror, arrive at the center ($\xi = 0$, $\eta = 0$) of the right-side mirror with the same phase, so that their fields are additive. The rays arriving at any point ($\xi \neq 0$, $\eta \neq 0$) of the right-side mirror from the points (x, y) and $(-x, -y)$ add up and also yield a real quantity; an in-phase current on the left-side mirror ensures an in-phase current also on the right-side mirror.

However, the reality of $w(x, y)$ is important also in another respect, namely in the analysis of some optimal properties of confocal resonators, to which we now proceed.

We have indicated in Item 1 that the phase correction (4.4) or (4.5) is optimal in the sense that it shapes a beam for which the radiation losses are the smallest possible at the given frequency, given beam cross section, and given distance between the coordinates. We now prove this premise.

Assume that the field $u(x, y)$ is specified on the aperture, that is, where the beam crosses the plane $z = 0$. According to (3.2), the field at the section $z = L$ is given in terms of u by the formula

$$v(\xi, \eta)=\int u(x, y)K(x, y, \xi, \eta)\,dx\,dy. \qquad (4.22)$$

The kernel K is given in (3.8); it satisfies the condition $K(x, y, \xi, \eta) = K(y, x, \eta, \xi)$. In (4.22) and in all the succeeding formulas of that Item, the integral has been taken over the area of the collector aperture, that is, over the section of the beam at $z = 0$ and $z = L$. We set up[43] the quantity

$$\mathscr{L}[u(x, y)] = \frac{\int v(\xi, \eta) v^*(\xi, \eta) d\xi d\eta}{\int u(x, y) u^*(x, y) dx dy} \tag{4.23}$$

and call it the transmission coefficient. Strictly speaking, $\mathscr{L}$ is the ratio not of the integral energy fluxes across the first and second apertures, but of the integral energy densities. The energy density and flux are proportional to each other only for fields close to plane waves. If $u(x,y)$ or $v(\xi, \eta)$ vary noticeably over distances of the order of the wavelength, then it is necessary to write for the energy transfer coefficient a different expression which, however, is much more difficult to investigate than (4.23). Assuming expression (4.23) for $\mathscr{L}$, we exclude from consideration "superdirectional" systems, analogous to superdirectional antennas.

The functions $u(x,y)$ obtained below, which guarantee a maximum of the functional $\mathscr{L}$, are smooth functions. Consequently, they really ensure an optimal solution among the non-superdirectional systems.

Substituting (4.22) in (4.23) we obtain

$$\mathscr{L}[u(x, y)] = \frac{\iint u(x, y) u^*(x', y') K_1(x, y, x', y') dx dy dx' dy'}{\int u(x, y) u^*(x, y) dx dy}, \tag{4.24}$$

where

$$K_1(x, y, x', y') = \int K(x, y, \xi, \eta) K^*(x', y', \xi, \eta) d\xi d\eta. \tag{4.25}$$

The functional (4.24) has an extremum, as is well known, if the function $u(x,y)$ satisfies the equation

$$\int u(x, y) K_1(x, y, x', y') dx dy = \Lambda u(x', y'). \tag{4.26}$$

Its eigenvalues are real, since the kernel K_1 (4.25) is hermitian; the largest of them is equal to the sought maximum of the coefficient $\mathscr{L}$[43].

Let us show, following[24], that the field of the natural waves of confocal resonators (or of corresponding lines) satisfies (4.26). The already noted reality of the functions $w(x,y)$ —the eigenfunctions of (3.10) for the phase correction (4.5) —signifies that the functions $u(x,y)$ satisfy the condition

$$u^*(x, y) = u(x, y) e^{-i\varphi(x, y)}. \tag{4.27}$$

Equation (3.7) has therefore for fields of confocal resonators the form

$$\int u(x, y) K(x, y, \xi, \eta) dx dy = \lambda u^*(\xi, \eta). \tag{4.28}$$

Let us take an equation which is the complex conjugate of (4.28) and form an iterated equation. We then obtain an equation identical with (4.26), the eigenvalue of which is the quantity $|\lambda|^2$.

Thus, the beam shaped by the correctors (4.5) is optimal, since the function (4.5) has the property that equation (3.10) generates real eigenfunctions. The largest attainable value of the transfer coefficient (4.23) is equal to the square of the modulus of the largest eigenvalue of (3.7) and (3.10).

5. It is obvious that the field formed in a line or in a resonator with optimal phase correctors ensures the smallest losses also for single transmission between two antennas with identical apertures, located at a relatively small distance of order ka^2 from one another. However, there exists also a simple and fruitful analogy[44] between transmission at short and long distances. It can be shown that the functions $w(x,y)$ obtained above solve simultaneously also the problem of antennas with optimal directivity patterns. It turns out that if an in-phase field is produced on the plane aperture of an antenna, with a distribution given by the solution of (3.10), then the radiation pattern of such an antenna ensures concentration of maximum relative energy within a specified angle 2α.* The parameter c for this distribution is equal to

$$c = ka\alpha. \tag{4.29}$$

It is known that the width of the directivity pattern is of the order of $2\pi/ka$. Equation (4.29) and the condition $c/2\pi \geq 1$ obtained above are necessary in order for the energy outside the given angle to be small, and show that if one does not go over to superdirectional antennas, the directivity pattern of the beam cannot be noticeably decreased. However, it is just the production of an optimal field distribution (according to Fig. 10) which enables us to make the summary energy outside the angle 2π quite small. This energy is equal to the radiation losses in the equivalent confocal system. For circular diaphragms it is given by the curve of Fig. 9; the parameter c is determined from (4.29). In order, for example, that the energy in the side lobes not to exceed 4×10^{-3} of the total radiated energy, the ratio of the radius of the aperture to the wavelength should be not less than $0.8/\alpha$. A small increase in c enables us to decrease the energy in the side lobes practically to zero, but it is impossible to reduce noticeably α for a specified ka, since Λ increases very rapidly with decreasing c.

Let us formulate once more the condition for the optimality of the re-radiators located in the near zone, and the condition for the optimality of the antennas. Assume that the dimension of the correctors, the wavelength, and the distance between correctors are specified. This defines the parameter c (3.1). By making the correctors confocal, we can ensure the maximum possible energy transfer between them; the losses depend on the value of c and are obtained from Figs. 8 and 9. This produces an in-phase field on the concave surface of the mirror (or on the plane surface in the middle of the lens). The distribution of the field over the radius for the symmetrical (most convenient) distribution for round correctors is given in Fig. 10. The

*For the two-dimensional problem this follows from a comparison of the results contained in[38,45-48]; however, a general proof valid also for a rectangular and round aperture can be given.

same distribution, produced on a plane surface, results in the far zone in an optimal diagram for a solid angle equal to the angle at which one corrector can be seen from the center of the neighboring one. The parameter c for the diagram is given by formula (4.29).

6. The calculation of systems which are under quasioptical conditions reduces formally to asymptotic diffraction problems. These problems have a character different from the classical problems, for which effective calculation methods have been developed long ago. The use of the parabolic equation has not led so far to qualitatively new results. A theory based on wave equations or Maxwell's equations is always complicated, and the transition to the asymptotic conditions is very cumbersome.

On the other hand, although direct application of the ray representations is not justified, it nevertheless seems to us that in the construction of a general and effective theory the starting point should be some modification of ray optics rather than wave optics. This modification should consist in the fact that within the concepts of geometrical optics there will be included several—two, three (not more)—of the simplest results of diffraction theory. It may be sufficient to use the theory of the focal spot and the theory of diffraction on a half-plane. Were such a simple extension of the ray treatment to be successful, one could attempt to apply to the quasioptical problems the entire tremendous arsenal of geometrical optics. This apparently is the direction in which[49] is oriented.

It would be desirable, for example, to formulate the ray treatment for confocal systems in such a way as to make rigorous precisely the geometrical-optic proof of their optimality. Then it would probably be possible to clarify the question of tolerances and the character of those serious disturbances occurring in the operation of confocal systems under certain deformations of the correctors[40-42]. The development of input converters for lines which match the field of the source with the field of the wave in the line should probably also be based on the optical theory of lenses. A more confident introduction of geometrical optics in the theory of broad waveguides will make it possible to construct a flexible and effective theory of prisms, lenses, and other wave guide devices.

Note added in proof. The parabolic equation was used not only in[24] but also in[50-53]. Of particular interest is the treatment in[50,52] of the exponential decrease of the field for small values of m (4.10) as the manifestation of caustics that bound ray bundles. In[51], which is related to[29], use was made in the derivation of (3.10) of the fact that following replacement of L by $x - \xi$ the factor in the kernel of (3.8), which depends on x and y, becomes the Green's function of the parabolic equation. In[35] it has been proposed to use the parabolic equation to refine the geometrical-optical calculation of lenses.

[1] A. E. Solomonovich, UFN 77, 589 (1962), Soviet Phys. Uspekhi 5, 625 (1963).

[2] J. Dees and A. Sheppard, Electronics **36**(2), 33 (1963).

[3] R. Fellers, Microwave J. **5**, 80 (1962).

[4] L. D. Rozenberg, Rasskaz o neslyshnom zvuke (Story of Inaudible Sound), AN SSSR, 1961.

[5] Yu. I. Kaznacheev, Shirokopolosnaya dal'naya svyaz' po volnovodam (Broadband Long-distance Communication over Waveguides), AN SSSR, 1959.

[6] S. Miller, PIRE **50**(2), 1215 (1962).

[7] P. L. Kapitza, Elektronika bol'shikh moshchnosteĭ (High-power Electronics), AN SSSR, 1962.

[8] Comte, Carfort, Pouthus, and Paris, Câbles et Trans. **11**(4), 342 (1957).

[9] Marcatili, BSTJ **40**(1), 185 (1961).

[10] R. B. Vaganov, Radiotekhnika i élektronika **8**, 1264 (1963).

[11] B. Z. Katsenelenbaum, Radiotekhnika i élektronika **8**, 1111 (1963). DAN SSSR **144**, 382 (1962).

[12] B. V. Kostrov and E. I. Nefedov, Radiotekhnika i élektronika **9**, no. 4 (1964).

[13] I. P. Kotik and A. N. Sivov, ibid. **9** (1964).

[14] L. A. Vaĭnshteĭn, Diffraktsiya élektromagnitnykh i zvukovykh voln na otkrytom kontse volnovoda (Diffraction of Electromagnetic and Sound Waves by the Open End of a Waveguide). Soviet Radio, 1953.

[15] B. Noble, Application of the Wiener-Hopf Method to the Solution of Partial Differential Equations (Russ. Transl.) IL, 1962.

[16] News Section, Radiotekhnika i élektronika **8**, 1087 (1963).

[17] G. Goubau and F. Schwering, IRE Trans. **AP-9**(3), 248 (1961).

[18] J. Christian and G. Goubau, IRE Trans. **AP-9**(3), 256 (1961).

[19] C. Tolbert and A. Straiton, PIRE **49**(3), 649 (1961).

[20] Electronic Design, No. 15, 18 (July 1961).

[21] L. Bertram, Microwave J. **5**(8), 69 (1961).

[22] G. D. Malyuzhinets, UFN **69**, 321 (1959), Soviet Phys. Uspekhi **2**, 749 (1960).

[23] B. Z. Katsenelenbaum, Author's Certificate (Patent) 21a4, 4868, No. 146362, of 26 May 1961; Byulleten' izobreteniĭ (Invention Bulletin) No. 8, 30 (1962).

[24] N. G. Bondarenko and V. I. Talanov, Abstracts of Papers at the Second All-union Symposium on Wave Diffraction, Gor'kiĭ, 7-13 June 1962. AN SSSR; Izv. vuzov (Radiofizika) (1964).

[25] B. Z. Katsenelenbaum, Radiotekhnika i élektronika **8**, 1516 (1963).

[26] A. I. Barchukov and A. M. Prokhorov, ibid. **4**, 2094 (1959).

[27] A. Fox and T. Li, BSTJ **40**(2), 453 (1961).

[28] A. Fox and T. Li, PIRE **48**(11), 1904 (1961).

[29] L. A. Vaĭnshteĭn, JETP **44**, 1050 (1963), Soviet Phys. JETP **17**, 709 (1963).

[30] Zimmerer, Anderson, Strine, and Beers, IEEE Trans., MTT-11, No. 2, 142 (1963).

[31] F. Schwering, Arch. d. Electr. übert. **15**(12), 555 (1961).

[32] G. Kotik and M. Newstein, J. Appl. Phys. **32**(2), 178 (1961).

[33] S. Morgan, IEEE Trans., MTT-11, No. 3, 191 (1963).

[34] J. Christian and G. Goubau, PIRE **49**(11), 1679 (1961).

[35] P. Connes, J. phys. et radium **19**, 262 (1958).

[36] A. P. Kirichenko, Izmerit. tekhn. (Measurement Engineering) No. 1, 18 (1963).

[37] I. S. Gradshteĭn and I. M. Ryzhik, Tablitsa integralov, summ, ryadov i proizvedeniĭ (Table of Integrals, Sums, Series, and Products), Fizmatgiz, 1962.

[38] G. Boyd and J. Gordon, BSTJ **40**(2), 489 (1961).

[39] J. Beyer and E. Scheibe, IRE Trans. **AP-10**, No. 3 (1962).

[40] G. Boyd and H. Kogelnik, BSTJ **41**(4), 1347 (1963).

[41] A. Fox and T. Li, PIEEE **51**(1), 80 (1963).

[42] R. Zimmerer, PIEEE **51**(1), 474 (1963).

[43] A. Kay, IRE Trans. **AP-8**(6), 586 (1960).

[44] M. Pokras, Periskopicheskie antenny i besprovodnye linii peredachi (Periscopic Antennas and Wireless Transmission Lines), Svyaz'izdat, 1963).

[45] M. S. Gurevich, Radiotekhnika i élektronika **1**, 313 (1956).

[46] D. Slepian and H. Pollak, BSTJ **40**(1), 43 (1961).

[47] Ya. I. Khurgin and V. P. Takovlev, Metody teorii tselykh funktsiĭ v radiofizike, teorii svyazi i optike (Methods of the Theory of Entire Functions in Radiophysics, Communication Theory, and Optics).

[48] H. Landau and H. Pollak, BSTJ **40**(1), 65 (1961).

[49] G. Goubau, Symposium on Electromagnetic Theory and Antennas, Copenhagen, June 1962, Proceedings, 907.

[50] L. A. Vaĭnshteĭn, JETP **45**, 684 (1963), Soviet Phys. JETP **18**, 471 (1964).

[51] L. A. Vaĭnshteĭn, ZhTF **34**, 193 (1964), Soviet Phys. Tech. Phys. **9**, 157 (1964).

[52] L. A. Vaĭnshteĭn, ZhTF **34**, 205 (1964), Soviet Phys. Tech. Phys. **9**, 166 (1964).

[53] B. Z. Katsenelenbaum, Radiotekhnika i élektronika **9** (1964).

Translated by J. G. Adashko
21