1 Introduction

1.1 The Origins of Sinusoidal Waves in Radio Transmission

Sinusoidal waves have become so universal in radio communications that few people are aware it was not always so. Heinrich Hertz (1893) used a spark discharge to produce the electromagnetic waves for his experiments. These waves would be called colored noise today. Spark gaps and arc discharges between carbon electrodes were the dominant wave generators for about 20 years after Hertz's experiments, and radio signals consisted of short or long bursts of colored noise. The development of rotating high-frequency generators and the electronic tube eventually made the generation of sinusoidal currents and waves possible.

A strong incentive to use sinusoidal waves was provided by the need to operate several transmitters at the same time but to receive them selectively. Maxwell[1] (1891) had already studied what we now call the resonance of a circuit with coil and capacitor. Many people worked on the theoretical investigation and the practical implementation of this phenomenon, but the credit for the introduction of resonating filters using coils and capacitors for the selective reception of radio signals is usually given to Marconi on the strength of his patents[2] (Marconi, 1901, 1904). Apparently, no one ever raised the question seriously whether sinusoidal waves were the only ones for which the phenomenon of resonance existed. Hence, transmitters and receivers were developed on the basis of sinusoidal waves. Regulation followed common practice and brought the assignment of frequency bands for various radio services. However, a quotation from a textbook published in 1920 shows that nonsinusoidal waves were still used at that time, and that this was fully understood[3]

---

[1] Paragraph 779, "Combination of the electrostatic capacity of a condenser with the electromagnetic capacity of self-induction of a coil."

[2] Patent 763 772 introduces the term *tuning* and describes how a transmitter and a receiver can be tuned by proper choice of inductance and capacitance. Patent 676 332, applied for seven months later, uses the term *resonance* in the description of what we would now call an *LC* parallel circuit. The patenting of resonance for selective radio transmission created one of the worst controversies in the controversy-rich history of early radio.

[3] Figure 22 on page 22 shows a nonsinusoidal carrier for telegraphy, which was typical for the spark transmitters of early radio transmission.

(Edelman, 1920, p. 187): "The more advanced methods of wireless communication utilize continuous [sinusoidal] waves, produced either by an arc, quenched spark, or direct high-frequency generator. Inasmuch as these methods are quite likely to be developed into the ultimate perfect wireless system, some consideration of the theory together with experimental operation is worthy of attention."

Let us turn to the mathematical basis for the phenomenon of resonance. The homogeneous differential equation

$$(d^2v/dt^2) + \omega_0^2 v = 0 \tag{1}$$

has the general solution

$$v(t) = V_1 \sin \omega_0 t + V_2 \cos \omega_0 t \tag{2}$$

Let a force function[1] $\omega_0^2 v_f(t)$ replace the zero on the right-hand side of Eq. (1):

$$(d^2v/dt^2) + \omega_0^2 v = \omega_0^2 v_f(t) \tag{3}$$

The general solution of this inhomogeneous equation consists of the general solution of the homogeneous equation, given by Eq. (2), plus a particular solution of the inhomogeneous equation. The systematic way to find such a particular solution is by means of the method of *variation of the constant* or the *Laplace transform*. However, in simple cases one usually tries to shorten these methods by guessing a particular solution.

Let $v_f(t)$ be a sinusoidal function:

$$v_f(t) = V \sin \omega t \tag{4}$$

We guess that a particular solution of Eq. (3) has the form

$$v_p(t) = V_0 \sin \omega t \tag{5}$$

Insertion of $v_p(t)$ for v, and of the force function of Eq. (4) for $v_f(t)$ in Eq. (3), yields the value of V_0:

$$V_0 = \frac{\omega_0^2}{\omega_0^2 - \omega^2} V, \qquad v_p(t) = \frac{\omega_0^2}{\omega_0^2 - \omega^2} V \sin \omega t \tag{6}$$

Evidently, our guessed particular solution holds for all values of ω except for $\omega = \omega_0$, which is called the *resonance case*. We have to guess a new solution for this case:

$$v_p(t) = V_0 \omega_0 t \cos \omega_0 t + V_1 \omega_0 t \sin \omega_0 t \tag{7}$$

---

[1] We assume that t has the dimension of time, ω_0 that of inverse time, and v that of voltage. By using $\omega_0^2 v_f(t)$ rather than $v_f(t)$ we assure that $v_f(t)$ also has the dimension of voltage.

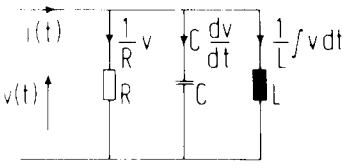

FIG. 1.1-1. Parallel resonant circuit for sinusoidal functions.

Insertion of $v_p(t)$ for v, and $V \sin \omega_0 t$ for $v_f(t)$, in Eq. (3) yields V_0 and V_1:

$$V_0 = -\tfrac{1}{2}V, \qquad V_1 = 0 \tag{8}$$

Hence, a particular solution for the resonance case equals

$$v_p(t) = -\tfrac{1}{2}V\omega_0 t \cos \omega_0 t \tag{9}$$

Let us now connect this purely mathematical concept of resonance with the resonance of a simple electrical circuit. Figure 1.1-1 shows a parallel resonant circuit with inductance L, capacitance C, and resistance R. The current $i(t)$ flowing into the circuit and the voltage $v(t)$ across the circuit are connected by the following differential equation:

$$\frac{1}{R}v + \frac{1}{L}\int v\,dt + C\frac{dv}{dt} = i(t) \tag{10}$$

Differentiation and reordering of the terms yields

$$\frac{d^2v}{dt^2} + \frac{1}{RC}\frac{dv}{dt} + \frac{1}{LC}v = \frac{1}{C}\frac{di}{dt} \tag{11}$$

The two equations (3) and (11) become equal if R is sufficiently large, and if the relations

$$\omega_0^2 = \frac{1}{LC}, \qquad v_f(t) = \frac{1}{\omega_0^2 C}\frac{di}{dt} = L\frac{di}{dt}$$

are satisfied. The mathematical concept of resonance of a differential equation is thus connected with the concept of resonance of an electronic circuit.

The differential equations considered so far were linear and had constant coefficients. Such differential equations will only resonate with sinusoidal force functions.[1] Furthermore, a lumped circuit with linear, time-invariant components is always described by a linear differential equation with constant coefficients, and will thus resonate with sinusoidal functions only. However, the mathematical concept of resonance is more gen-

[1] More precisely, they will resonate only with the functions $\exp(-st)$, $\exp(-st) \sin \omega t$, and $\exp(-st) \cos \omega t$, where s is a nonnegative real number.

eral. In particular, it applies to linear differential equations with variable coefficients. They describe lumped electronic circuits with linear, time-*variable* components. Examples of such linear, time-variable components are switches, microphones, and modulators.[1] In the early days of radio communications there were no good time-variable components, and the phenomenon of resonance could thus be exploited for sinusoidal functions only. This situation was decisively changed by the advent of semiconductor technology. The switch is now one of the most desirable electronic components, and it is a linear, yet time-variable component. Let us then investigate some lumped circuits with linear, time-variable components that resonate with nonsinusoidal functions.

1.2 Lumped, Time-Variable Resonant Circuits

In order to derive resonant circuits for nonsinusoidal functions we redesign first the parallel resonant circuit of Fig. 1.1-1 for implementation by operational amplifiers. To this end we integrate Eq. (1.1-11) twice,

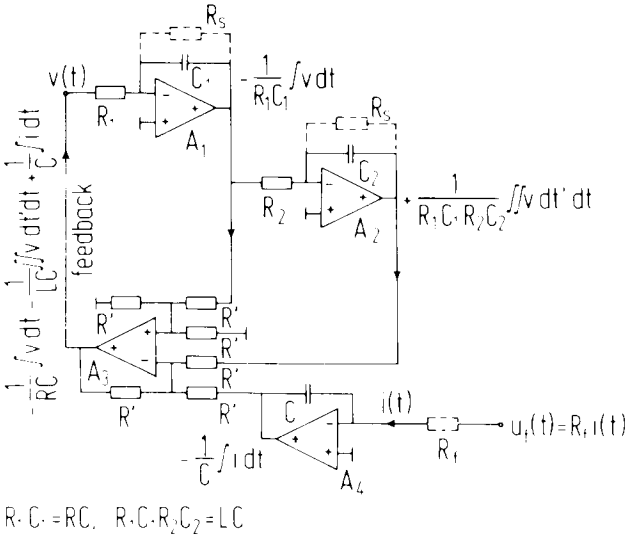

FIG. 1.2-1. Implementation of the parallel resonant circuit of Fig. 1.1-1 by means of operational amplifiers, resistors, and capacitors.

[1] Many textbooks call amplitude or frequency modulators nonlinear devices, and go on to discuss their use for the transmission of voice and music. There are two causes for this common error. First, no distinction is made between *linear*, *time-variable*, and *nonlinear*. Second, it is overlooked that a circuit with nonlinear components may have a linear relationship between input and output. For instance, a ring modulator with four square-law rectifiers transforms two input voltages u and v into the output voltage $(u + v)^2 - (u - v)^2 = 4uv$. The nonlinear terms u^2 and v^2 are canceled, and only the bilinear term uv remains.

$$v + \int \frac{1}{RC} v\, dt + \frac{1}{LC} \int\int v\, dt'\, dt = \frac{1}{C} \int i\, dt \tag{1}$$

and rearrange the terms as follows:

$$v = -\frac{1}{RC} \int v\, dt - \frac{1}{LC} \int\int v\, dt'\, dt + \frac{1}{C} \int i\, dt \tag{2}$$

Consider now the circuit of Fig. 1.2-1. The voltage $v(t) = v$ is fed in the upper left corner to the integrator[1] composed of the operational amplifier A_1, the resistor R_1, and the capacitor C_1. The integrated voltage $-(1/R_1C_1)\int v\, dt$ is produced. A second integration by A_2, R_2, and C_2 yields the voltage $+(1/R_1C_1R_2C_2)\int\int v\, dt'\, dt$. Furthermore, the current $i(t)$ in the lower right-hand corner is integrated by the amplifier A_4 and the capacitor C to yield the voltage $-(1/C)\int i\, dt$. These three voltages are summed with proper signs to yield

$$-\frac{1}{R_1C_1} \int v\, dt - \frac{1}{R_1C_1R_2C_2} \int\int v\, dt'\, dt + \frac{1}{C} \int i\, dt$$

If one chooses

$$R_1C_1 = RC, \qquad R_1C_1R_2C_2 = LC \tag{3}$$

one obtains just the right-hand side of Eq. (2). Since the right-hand side equals the voltage $v = v(t)$ on the left-hand side, we may close the feedback loop in Fig. 1.2-1. Hence, the circuit of Fig. 1.2-1 is a practical implementation[2] of Eq. (2).

The circuits of Figs. 1.1-1 and 1.2-1 will resonate with "periodic" sinusoidal functions, or functions that have sufficiently many periods. As the first generalization of such resonating circuits we will discuss one that resonates with sinusoidal or cosinusoidal *pulses* that have $i = 1, 2, \ldots$ periods and are zero outside the interval $-T/2 \leqq t \leqq +T/2$. The first few of these pulses are shown in Fig. 1.2-2, together with a rectangular pulse. These are the functions used for the Fourier series in a finite interval. They are defined by the equations

$$\begin{aligned} (d^2v/dt^2) + \omega^2 v &= 0, \qquad -T/2 \leqq t \leqq +T/2 \\ v &= 0, \qquad t < -T/2, \quad t > T/2 \end{aligned} \tag{4}$$

---

[1] It is assumed that the reader is familiar with the use of operational amplifiers for integration, summation, and differentiation. An excellent text for this field is the book by Graeme *et al.* (1971).

[2] The resistors R_s shown by dashed lines prevent the operational amplifiers from saturating due to drift. We do not discuss here the many technical refinements used in practical circuits with operational amplifiers, since we are interested in principles.

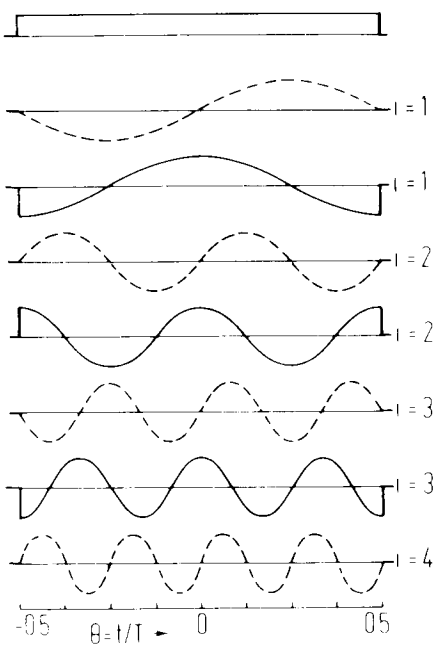

FIG. 1.2-2. The first eight functions of the Fourier series consisting of a constant, the sine functions $\sqrt{2}\sin 2\pi it/T$, and the cosine functions $\sqrt{2}\cos 2\pi it/T$.

We integrate the differential equation twice and add a force function $v_f(t)$ on the right-hand side:

$$v + \omega^2 \int\int v\, dt'\, dt = v_f(t) \tag{5}$$

The terms are rearranged to make v the only term on the left-hand side:

$$v = -\omega^2 \int\int v\, dt'\, dt + v_f(t) \tag{6}$$

Figure 1.2-3 shows the voltage $v(t)$ applied to the integrator consisting of amplifier A_1, resistor R_1, and capacitor C_1. The output voltage $-(1/R_1C_1)\int v(t)\, dt$ is integrated a second time to yield $+(1/R_1C_1R_2C_2)\int\int v\, dt'\, dt$. This voltage is fed through a multiplier M that multiplies with $+1$ during the time interval $-T/2 \leqq t \leqq +T/2$ and with 0 otherwise. The summation of this voltage with a force function $v_f(t)$ according to Fig. 1.2-2 yields

$$-\omega^2 \int\int v\, dt'\, dt + v_f(t)$$

in the interval $-T/2 \leqq t \leqq +T/2$ and zero otherwise. Closing of the feedback loop thus produces a circuit that implements Eq. (4).

In Fig. 1.2-1 we used the two resistors R_s to prevent saturation of the integrators due to drift. In Fig. 1.2-3 we use instead reset switches S_r that are always closed except during the interval $-T/2 \leqq t \leqq +T/2$. These

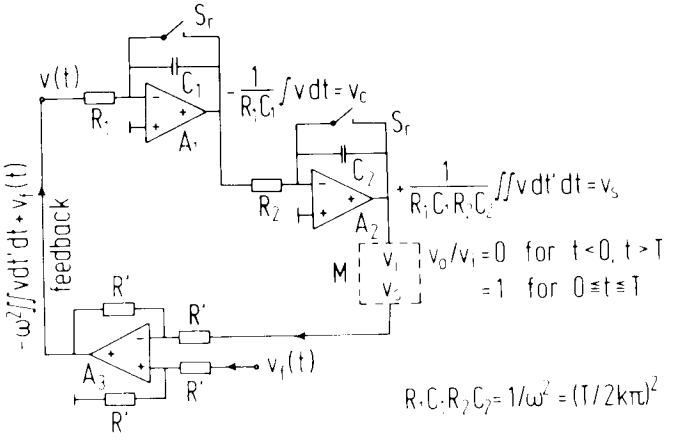

FIG. 1.2-3. Resonant circuit for the functions of Fig. 1.2-2. The choice of the time constant $R_1C_1R_2C_2$ determines the number $k = 1, 2, \ldots$ of the cycles of the sine or cosine pulse to which the circuit is tuned. A resonating *cosine* pulse with $i = k$ produces a large voltage v_c, and $v_s = 0$; a resonating *sine* pulse with $i = k$ produces a large voltage v_s, and $v_c = 0$. The reset switches S_r are closed at $t = \pm T/2$.

switches make the output voltages of both integrators zero outside this interval, which duplicates the effect of the multiplier M. Hence, the multiplier may be left out.

The voltages produced at the output terminals of the two integrators in response to a single cycle ($i = 1$) of a sinusoidal or cosinusoidal force function $v_f(t)$ is shown in Fig. 1.2-4. One may readily see that this circuit discriminates between sine and cosine pulses with the same period $2\pi/\omega$. The sinusoidal pulse (a) and the cosinusoidal pulse (d) were produced by digital circuits, which explains their steps.

Figure 1.2-5 shows oscillograms of the output voltage v_c of the first integrator if the circuit is tuned to $k = 128$ by the choice of the product $R_1C_1R_2C_2 = (T/2k\pi)^2$, whereas the force function $v_f = V \cos 2\pi it/T$ with $i = 128$, 129, and 130 is applied; this means that the circuit is tuned for the detection of a cosine (output v_c) or a sine pulse (output v_s) of 128 cycles, and that cosine pulses with 128, 129, or 130 cycles are fed to its input.

We obtained Eq. (5) by integration of Eq. (4) and addition of the force function $v_f(t)$. Another method is to make the following substitutions:

$$d^2v/dt^2 = u, \qquad dv/dt = \int u\, dt, \qquad v = \int\int u\, dt'\, dt \tag{7}$$

One obtains from Eq. (4)

$$u + \int\int u\, dt'\, dt = 0 \tag{8}$$

A force function

$$u_f(t) = d^2v_f/dt^2 \tag{9}$$

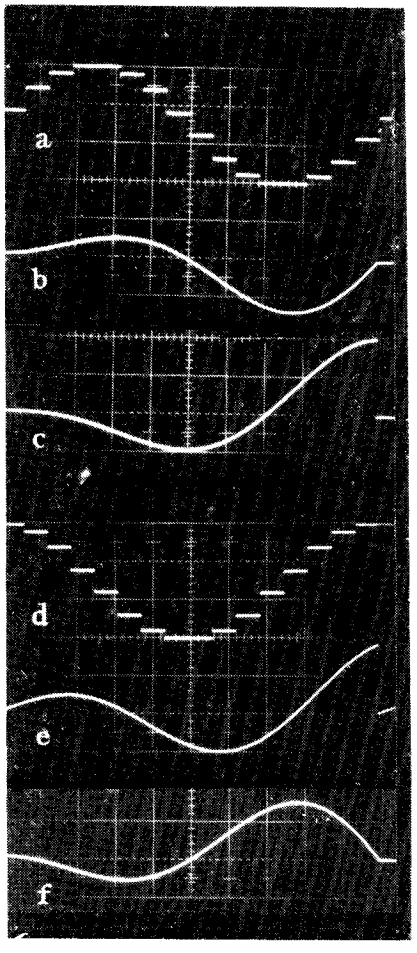

FIG. 1.2-4. Typical voltages of the circuit of Fig. 1.2-3: (a) force function $v_f(t) = V \sin 2\pi t/T$; (b) and (c) resulting voltages v_c and v_s; (d) force function $v_f(t) = V \cos 2\pi t/T$; (e) and (f) resulting voltages v_c and v_s. Horizontal scale: 15 msec/div. (Courtesy P. Schmid, R. Durisch, and D. Novak of Allen-Bradley Co., Milwaukee, Wisconsin.)

may be added on the right. The resulting equation is equal to Eq. (6), since the letters v are merely replaced by the letters u. The force function $u_f(t)$ in Eq. (9) differs only in sign from the force function $v_f(t)$ if $v_f(t)$ is a sinusoidal or cosinusoidal function.

The equality between integration and the substitutions according to Eq. (7) does not hold for differential equations with variable coefficients, to which we turn now. Figure 1.2-6 shows Legendre *pulses*. These pulses equal the Legendre polynomials $P_j(\theta)$ in the interval $-1 \leqq \theta \leqq +1$ and they are zero outside:

$$P_0(\theta) = 1, \qquad P_1(\theta) = \theta, \qquad P_2(\theta) = \tfrac{1}{2}(3\theta^2 - 1)$$

$$P_3(\theta) = \tfrac{1}{2}(5\theta^3 - 3\theta), \qquad P_4(\theta) = \tfrac{1}{8}(35\theta^4 - 30\theta^2 + 3) \tag{10}$$

$$-1 \leqq \theta \leqq +1, \qquad \theta = t/T$$

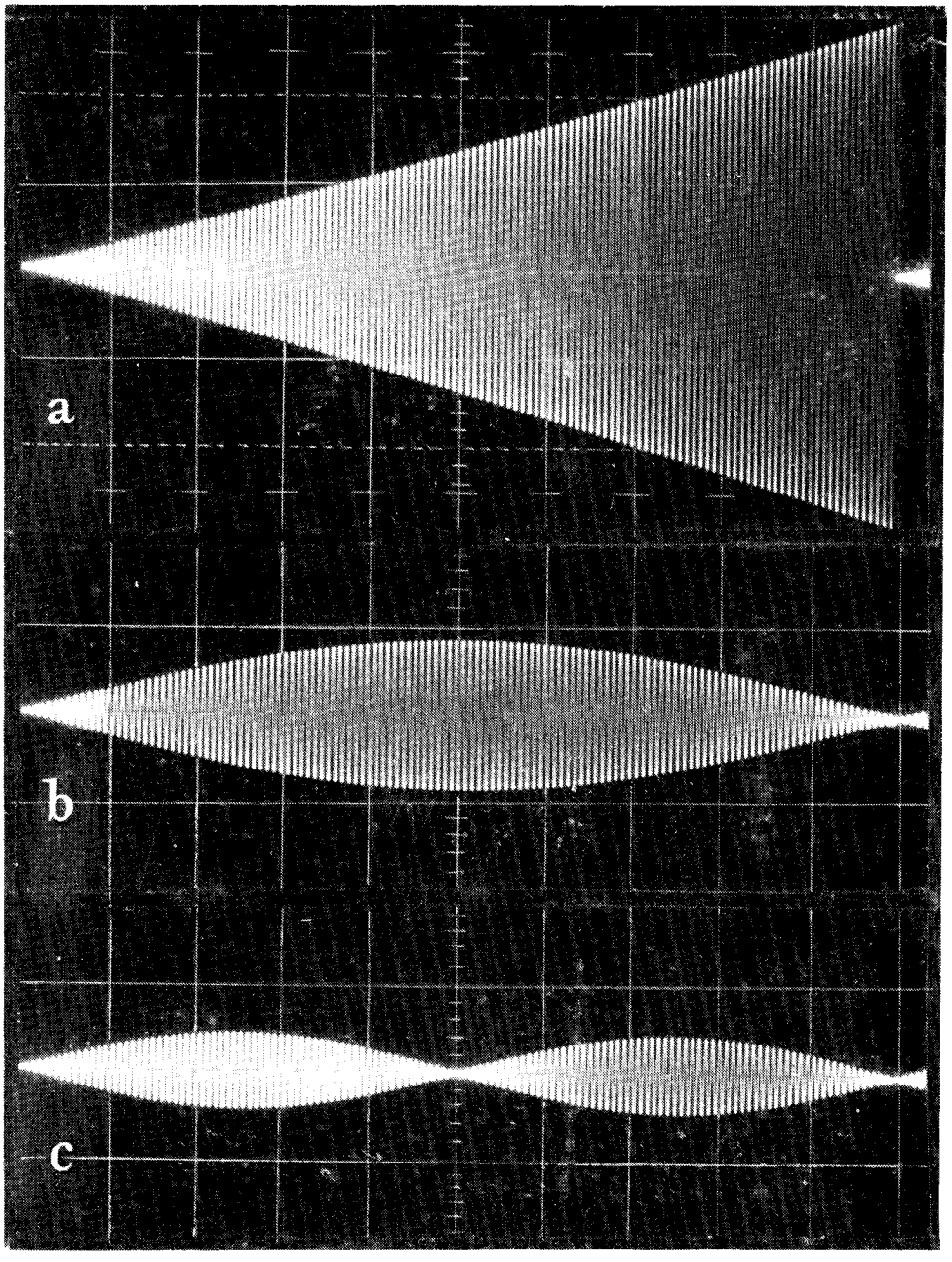

FIG. 1.2-5. Typical voltages of the circuit of Fig. 1.2-3. Circuit is tuned to resonate with sine and cosine pulses with $k = 128$ cycles. Output voltages v_c shown are caused by force functions $v_f(t) = V \cos 2\pi it/T$ with $i = 128$ (a), $i = 129$ (b), and $i = 130$ (c). Duration of the traces is $T = 78$ ms. (Courtesy P. Schmid, R. Durisch, and D. Novak of Allen-Bradley Co., Milwaukee, Wisconsin.)

The Legendre polynomials are defined[1] by the following differential equation:

$$(1 - \theta^2)\frac{d^2v}{d\theta^2} - 2\theta\frac{dv}{d\theta} + j(j + 1)v = 0, \qquad -1 \leqq \theta \leqq +1 \qquad (11)$$

[1] The mathematician says that the Legendre polynomials are the eigenfunctions of this differential equation, just as sinusoidal functions are the eigenfunctions of Eq. (4).

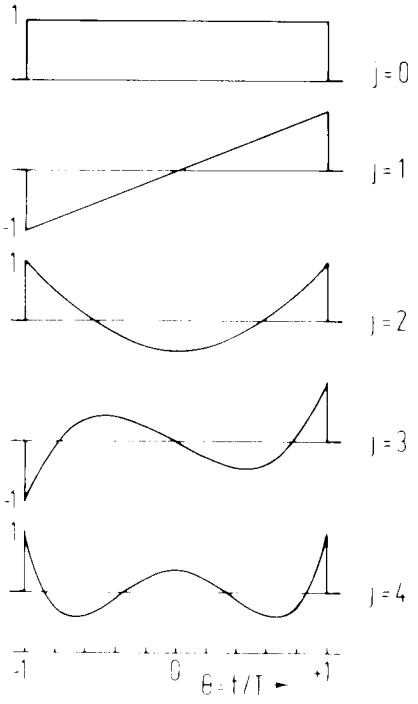

FIG. 1.2-6. Pulses of the form of the Legendre polynomials $P_j(\theta)$ in the interval $-1 \leqq \theta \leqq 1$ and zero outside.

We want equations that show the real time t rather than the normalized time $\theta = t/T$, and we thus substitute t/T for θ in Eq. (11):

$$(T^2 - t^2)\frac{d^2v}{dt^2} - 2t\frac{dv}{dt} + j(j+1)v = 0 \tag{12}$$

We use the substitutions of Eq. (7) and add a force function $u_f(t)$:

$$u - \frac{2t}{T^2 - t^2}\int u\,dt + \frac{j(j+1)}{T^2 - t^2}\int\int u\,dt'\,dt = u_f(t) \tag{13}$$

Next we rearrange the terms so that only u stands on the left-hand side of the equation, and so that the time constant $1/T$ appears conspicuously with every integration:

$$u = 2\frac{tT}{T^2 - t^2}\frac{1}{T}\int u\,dt - j(j+1)\frac{T^2}{T^2 - t^2}\frac{1}{T^2}\int\int u\,dt'\,dt + u_f(t) \tag{14}$$

A circuit implementing this equation is shown in Fig. 1.2-7. The voltage $u(t)$ is integrated twice to yield $\int u\,dt$ and $\int\int u\,dt'\,dt$. These voltages are multiplied by the multipliers M_1 and M_2 with $tT/(T^2 - t^2)$ and $T^2/(T^2 - t^2)$. Further multiplications by 2 and by $j(j+1)$ are accomplished by the amplifiers A_3, A_4, and the resistors associated with them.[1] The resulting

[1] These two multiplication circuits can be combined with the amplifier A_5.

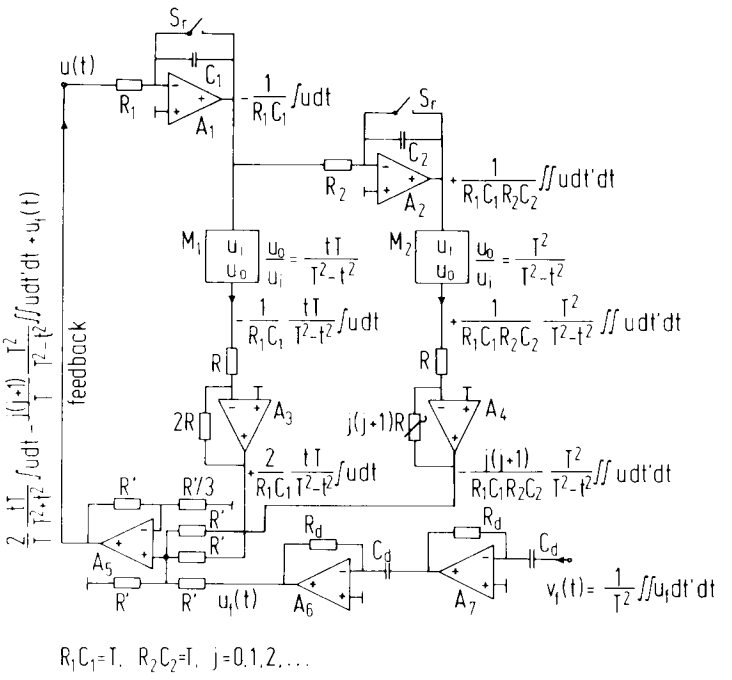

FIG. 1.2-7. Resonant circuit for the Legendre pulses of Fig. 1.2-6. The choice of the resistance $j(j+1)R$ in the feedback loop of the amplifier A_4 determines for which function $P_j(\theta)$ resonance occurs. The time constants R_1C_1 and R_2C_2 determine the time interval T of $\theta = t/T$.

voltages and the force function $u_f(t)$ are summed by the amplifier A_5, with attention to the proper positive or negative sign. The output voltage of A_5 equals the right-hand side of Eq. (14), and this voltage may thus be fed back to yield the input voltage $u(t)$ for the amplifier A_1.

The multipliers M_1 and M_2 have to produce the output voltages u_o from the input voltages u_i as functions of time. The ratios u_o/u_i as functions of t are shown in Fig. 1.2-8. In the days of the analog computer it was usual to implement such multipliers either by special, motor-driven potentiometers or by resistor networks and many switches. The modern way is to produce discrete values of the functions $tT/(T^2 - t^2)$ and $T^2/(T^2 - t^2)$ by means of a microprocessor, and feed the digital numbers to a digital–analog multiplier (Harmuth, 1979a, pp. 111–116).

Let us observe that circuits like the one in Fig. 1.2-7 were once routinely assembled for the simulation of differential equations by analog computers, but the advance of digital computers has made this almost a forgotten art (Korn and Korn, 1964; Johnson, 1963).

If the circuit of Fig. 1.2-7 is to resonate with Legendre pulses, we must feed the force function $v_f(t) = VP_j(t/T)$ into the circuit. However, our circuit implements the integral equation (14) for $u(t)$ rather than the differential equation (12) for $v(t)$. The force function $u_f(t)$ required to make it resonate is the second derivative of $v_f(t)$ according to Eq. (9). The circuit of

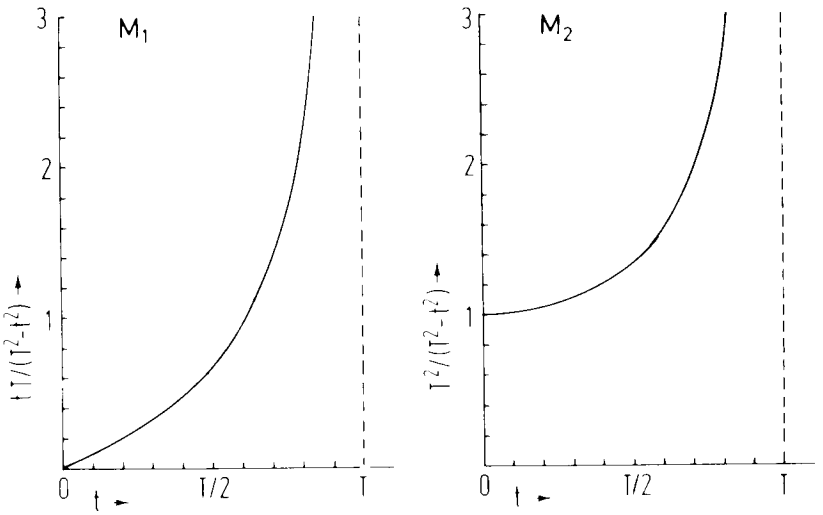

FIG. 1.2-8. Time variation of the ratio u_o/u_i of the two multipliers M_1 and M_2 in Fig. 1.2-7.

Fig. 1.2-7 shows at the bottom the voltage $v_f(t)$ fed through two differentiating stages (A_6, A_7) to yield $u_f(t)$. Such differentiations are undesirable. There are ways to avoid them, but we want to show here only how resonating circuits for nonsinusoidal functions can be devised, without advocating their construction and use.[1]

As the last example let us design a resonant circuit for Bessel functions. Figure 1.2-9 shows the first five functions $J_n(\theta)$. The functions $J_n(\theta)$ are defined by the following differential equation:

$$\theta^2 \frac{d^2v}{d\theta^2} + \theta \frac{dv}{d\theta} + (\theta^2 - n^2)v = 0, \qquad 0 \leqq \theta < \infty, \quad n = 0, 1, 2, \ldots \tag{15}$$

We replace θ by t/T, make the substitution of Eq. (7), add a force function $u_f(t)$, and rearrange the terms in analogy to Eq. (14):

$$u = -\frac{T}{t}\frac{1}{T}\int u\,dt - \left(1 - n^2\frac{T^2}{t^2}\right)\frac{1}{T^2}\int\int u\,dt'\,dt + u_f(t) \tag{16}$$

Figure 1.2-10 shows the circuit implementing the equation. The voltage $u(t)$ is integrated to yield $\int u\,dt$ and $\int\int u\,dt'\,dt$. The integrated voltages are multiplied with T/t and $1 - n^2T^2/t^2$ in the multipliers M_1 and M_2. The multiplied voltages and the force function $u_f(t)$ are summed by amplifier A_4, and the sum is fed back to amplifier A_1. The ratios u_o/u_i of the multipliers M_1 and M_2 are shown in Fig. 1.2-11. Note that the multiplier M_2 has

[1] One way is to find an integral equation rather than a differential equation that defines Legendre polynomials. Another is to use the circuit not for Legendre pulses, but for pulses obtained from the Legendre pulses by differentiating twice, e.g., $P_2''(\theta) = 3$, $P_3''(\theta) = 15\theta$, $P_4''(\theta) = (420\theta^2 - 60)/8$. More technical remedies are discussed by Graeme *et al.* (1971).

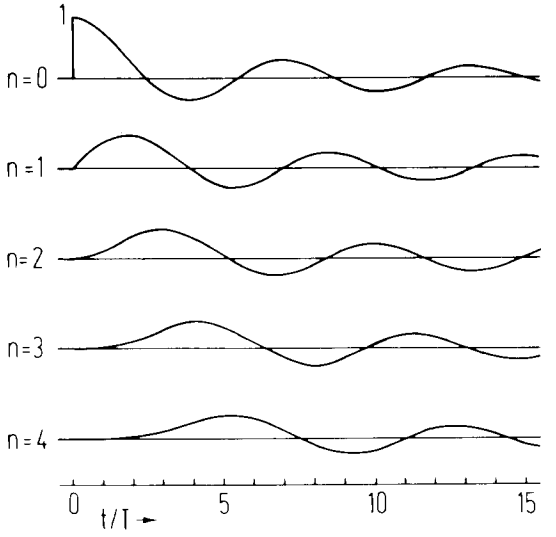

FIG. 1.2-9. Pulses of the form of the Bessel functions $J_n(\theta)$ in the interval $0 \leqq \theta < \infty$ and zero for $\theta < 0$.

to be changed for every value of $n = 0, 1, 2, \ldots$, while in Fig. 1.2-7 only a resistor needed changing for every value $j = 0, 1, 2, \ldots$. This causes no difficulty if the function $1 - n^2T^2/t^2$ is produced by a microprocessor, but it was a problem in bygone days when potentiometers and resistor networks had to be used.

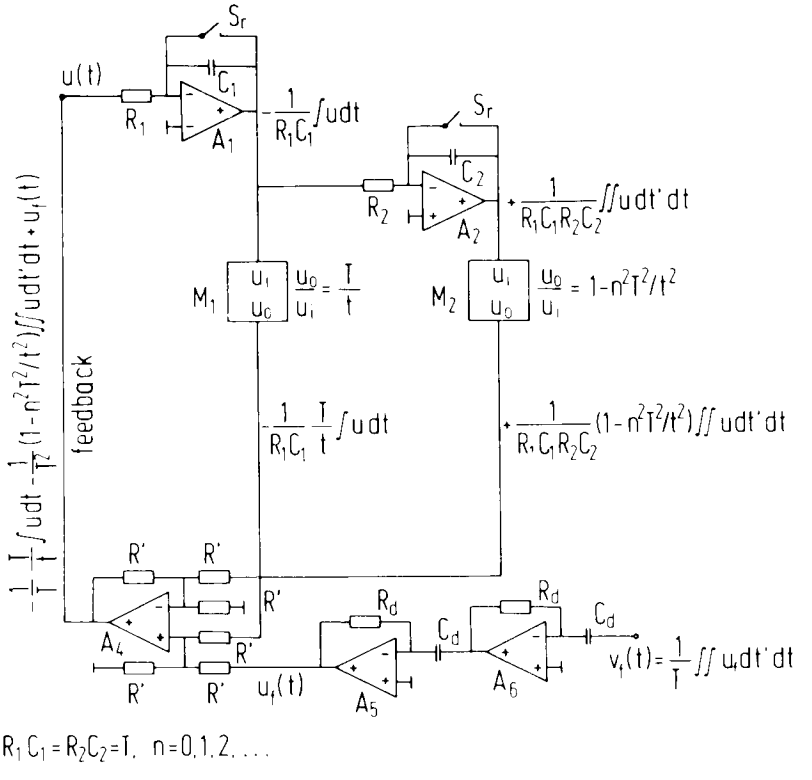

FIG. 1.2-10. Resonant circuit for the Bessel functions of Fig. 1.2-9. The choice of the ratio $u_0/u_i = 1 - n^2T^2/t^2$ of the multiplier M_2 determines for which function $J_n(\theta)$ resonance occurs. The time constants R_1C_1 and R_2C_2 determine the time interval T of $\theta = t/T$.

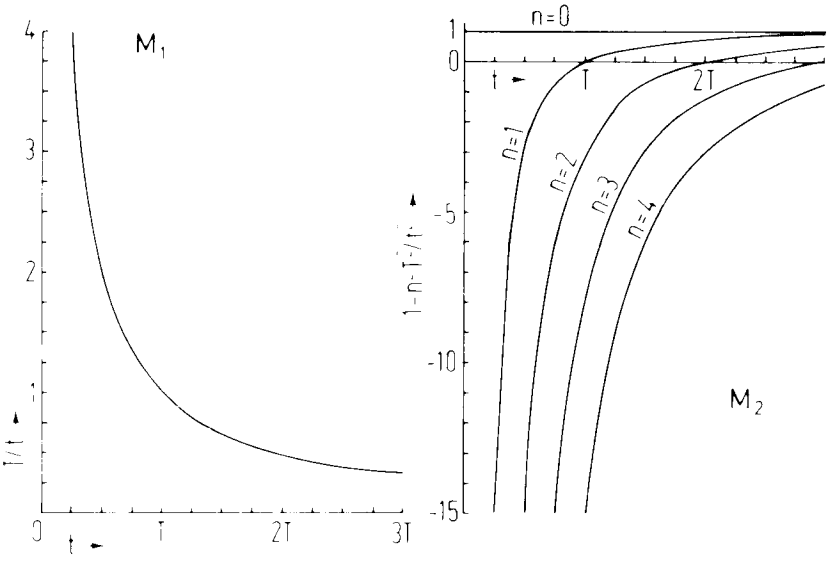

FIG. 1.2-11. Time variation of the ratio u_0/u_i of the two multipliers M_1 and M_2 in Fig. 1.2-10.

These examples should suffice to show that lumped resonant circuits are not restricted to sinusoidal functions. The examples also show that it is not easy, even today, to find resonant circuits that are as practical as the ones for sinusoidal functions. The circuits to be discussed later on will always be based on switches, sampled functions, and digital technology.

1.3 DISTRIBUTED, TIME-INVARIANT RESONANT CIRCUITS

A lumped circuit with linear, time-invariant components can only resonate with sinusoidal functions, while resonance with any other functions requires time-variable or nonlinear components. The reason is that lumped circuits are described by *ordinary* differential equations.

The same does *not* hold for distributed circuits and structures that are described by *partial* differential equations. To elaborate this statement we will study transmission lines, and derive the basic resonating circuit with time-invariant components for nonsinusoidal functions from them.

A transmission line is described by the following pair of partial differential equations:

$$\frac{\partial v}{\partial x} + L\frac{\partial i}{\partial t} + Ri = 0 \tag{1}$$

$$\frac{\partial i}{\partial x} + C\frac{\partial v}{\partial t} + Gv = 0 \tag{2}$$

The constants L, C, R, and G are the inductance, capacitance, resistance,

and conductance per unit length of the line, while $v = v(x, t)$ and $i = i(x, t)$ are the voltage and current at the location x at the time t.

In order to separate the variables v and i one may introduce a new function:

$$v = \partial w/\partial x \tag{3}$$

Equation (2) then yields

$$\frac{\partial}{\partial x}\left(i + C\frac{\partial w}{\partial t} + Gw\right) = 0 \tag{4}$$

from which follows

$$i + C\,\partial w/\partial t + Gw = K \tag{5}$$

where K is independent of x. Without loss of generality one may choose $K = 0$, since $v = \partial w/\partial x$ remains unchanged if one adds to w a term that is independent of x. Hence, one obtains the following relation between i and w:

$$i = -C\,\partial w/\partial t - Gw \tag{6}$$

Substitution of Eqs. (3) and (6) into Eq. (1) yields the telegrapher's equation:

$$\frac{\partial^2 w}{\partial x^2} - LC\frac{\partial^2 w}{\partial t^2} - (LG + RC)\frac{\partial w}{\partial t} - RGw = 0 \tag{7}$$

To obtain its general solution one introduces a new function $u(x, t)$,

$$w(x, t) = e^{-at}u(x, t) \tag{8}$$

where a is a constant. Substitution of Eq. (8) into Eq. (7) yields

$$\frac{\partial^2 u}{\partial x^2} - LC\left(a^2u - 2a\frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial t^2}\right) - (LG + RC)\left(-au + \frac{\partial u}{\partial t}\right) - RGu = 0 \tag{9}$$

If one chooses

$$a = (LG + RC)/2LC \tag{10}$$

one eliminates the terms $\partial u/\partial t$ in Eq. (9) and obtains a simpler equation:

$$\frac{\partial^2 u}{\partial x^2} - LC\frac{\partial^2 u}{\partial t^2} + \frac{1}{2}(LG - RC)u = 0 \tag{11}$$

The term $(LG - RC)u$ vanishes if the condition

$$LG - RC = 0 \tag{12}$$

is satisfied. This is the case of the *distortion-free transmission line*. The product RC is generally much larger than LG for practical transmission lines. Thus in order to satisfy Eq. (12) one must either increase the inductance L, e.g., by means of materials with large magnetic permeability, or the conductance G, by using low-quality insulation. The increase of the inductance was common before the introduction of coaxial cables, whereas the increase of the conductance was never practical.[1] However, the length of the transmission lines we are interested in here is of the order of meters rather than kilometers, and we can well afford to increase the conductance G by means of poor insulation so that the condition for distortion-free transmission is satisfied. Equation (11) is in this case reduced to the wave equation:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \qquad c^2 = \frac{1}{LC} \tag{13}$$

Its general solution was found by d'Alembert in the eighteenth century:

$$u(x, t) = g_1(x - ct) + g_2(x + ct) \tag{14}$$

Substitution into Eq. (8) yields

$$w(x, t) = e^{-at}[g_1(x - ct) + g_2(x + ct)] \tag{15}$$

Using Eqs. (3) and (6), and writing f and g for the derivatives g_1' and g_2' with respect to their arguments $x - ct$ or $x + ct$, yields the voltage $v(x, t)$ and the current $i(x, t)$:

$$v(x, t) = e^{-at}[f(x - ct) + g(x + ct)] \tag{16}$$

$$i(x, t) = (1/Z)e^{-at}[f(x - ct) - g(x + ct)] \tag{17}$$

$$Z = (L/C)^{1/2}, \quad c = (LC)^{-1/2}, \quad a = (LG + RC)/2LC, \quad LG - RC = 0$$

[1] Communication links must always be "practically" distortion-free, or one could not transmit information. Radio links in vacuum are inherently distortion-free, whereas cable links are made so by means of compensating circuits, called *equalizers*, that are inserted at certain intervals. Power transmission lines, on the other hand, do not need to transmit information, and the distortion problem is solved by using sinusoidal currents and voltages rather than equalizers. One cannot use this expedient for communication links since information is always transmitted by nonsinusoidal currents, voltages, or field strengths. The distinction between power transmission and information transmission was not made for almost a century of development of electrical communications. Even today, the education of the electrical engineer is based on a double-think: He is presented a theory based on sinusoidal functions, and then told that sinusoidal functions transmit information at the rate zero; at this point he must switch to think in terms of nondenumerably many sinusoidal functions, and he is left to figure out which parts of the learned theory apply to this abstraction. A paper by Hartley (1928) marks the approximate time when the difference between power transmission and information transmission became more clearly recognized.

The function $f(x - ct)$ propagates with time in the direction of larger values of x, whereas the function $g(x + ct)$ propagates in the opposite direction. Let us consider the function $f(x - ct)$ alone. If it started propagating at $x = 0$ at the time $t = 0$ it will have reached the point $x = ct$ at the time $t = x/c$, and we may replace at by ax/c:

$$v(x, t) = e^{-ax/c} f(x - ct) \tag{18}$$

$$i(x, t) = (1/Z)e^{-ax/c} f(x - ct) \tag{19}$$

Consider the circuit of Fig. 1.3-1. It consists of a hybrid coupler HYC1, an amplifier AMP, a second hybrid coupler HYC2, and a transmission line DEL of length L and delay T. This transmission line is shown as a coaxial cable, but Eqs. (1) and (2) apply to many other types of transmission lines as well.

Let the amplification be set so that the feedback loop in Fig. 1.3-1 has essentially unit gain. Then let a periodic signal with period T,

$$f(t) = f(t - mT), \quad m = 0, 1, 2, \ldots \tag{20}$$

be fed to the input terminal of the hybrid coupler HYC1. It travels through the feedback loop and arrives after a delay time T at the feedback terminal of HYC1 as the signal $f(t)$. At this time the signal $f(t - T) = f(t)$ is fed again to the input terminal, and the sum of the new signal with the feedback signal produces the signal $f(t) + f(t - T) = 2f(t)$ with twice the amplitude. After n periods of the signal one has the signal $nf(t)$ with n times the amplitude. This is as good a resonance effect as the one shown in Fig. 1.2-5a, where the amplitude of the oscillations also increases propor-

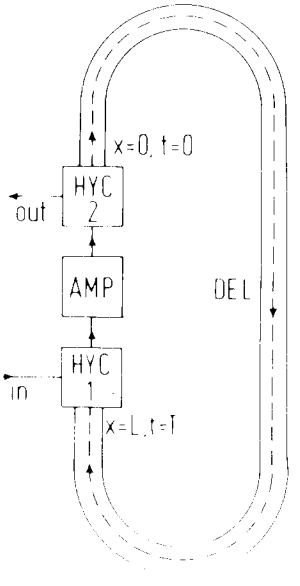

FIG. 1.3-1. Basic resonant circuit for general periodic functions with period T. HYC, hybrid coupler; AMP, amplifier; DEL, distortion-free delay line of length L and delay T.

tionate with the number of received periods of the sinusoidal signal. The difference is, of course, that any periodic function $f(t)$ will resonate with the distributed circuit of Fig. 1.3-1, whereas only sinusoidal functions will resonate with the lumped circuit of Figs. 1.2-1 and 1.2-3. Note that the circuit of Fig. 1.3-1 is linear and time-invariant. This circuit is as basic for nonsinusoidal waves as the LC circuit is for sinusoidal waves.

We have taken here great care to find the general solution for the distortion-free transmission line described by Eqs. (1), (2), and (12). Now we show that a completely different result is obtained by being not quite so careful. A typical method to solve a partial differential equation is to separate the variables by Bernoulli's product method. We apply this method to the telegrapher's equation (7):

$$w(x, t) = \varphi(x)\psi(t) \tag{21}$$

Substitution of $\varphi(x)\psi(t)$ into Eq. (7) yields two ordinary differential equations:

$$\begin{gathered} \frac{d^2\varphi}{dx^2} + (\mu - RG)\varphi = 0 \\ LC\frac{d^2\psi}{dt^2} + (LG + RC)\frac{d\psi}{dt} + \mu\psi = 0 \end{gathered} \tag{22}$$

Their general solutions are:

$$\varphi(x) = A_1 \cos(\mu - RG)^{1/2}x + A_2 \sin(\mu - RG)^{1/2}x \tag{23}$$

$$\psi(t) = B_1 e^{\gamma t} + B_2 e^{\delta t} \tag{24}$$

$$\begin{aligned} \gamma &= -a + (a^2 - \mu c^2)^{1/2}, & \delta &= -a - (a^2 - \mu c^2)^{1/2} \\ a &= (LG + RG)/2LC, & c &= (LC)^{-1/2} \end{aligned} \tag{25}$$

where μ is the eigenvalue to be determined by boundary conditions.

The function $w(x, t)$ now consists of sinusoidal and cosinusoidal terms multipled with an exponential function, e.g., $e^{\gamma t}\cos(\mu - RG)^{1/2}x$. The differentiation required to obtain the voltage v and the current i from $w(x, t)$ according to Eqs. (3) and (6) will produce additional terms of the same form. Hence, it appears that sinusoidal and exponential functions are distinguished by the telegrapher's equation. This is quite wrong. Bernoulli's product of Eq. (21) only yields particular solutions of the telegrapher's equation, not the general solution, and it is the method of solution rather than the equation that distinguishes the sinusoidal and exponential functions.

This is a very important point since the theory of electricity in general and the theory of electromagnetic radiation in particular—to the extent

used in this book—are derived from Maxwell's equations. Most of the known solutions of Maxwell's equations use Bernoulli's product, and they must all be avoided if one wants to use nonsinusoidal signals or signals with a large relative bandwidth.

1.4 RELATIVE BANDWIDTH

The relative bandwidth is usually defined as a quotient bandwidth/(carrier frequency). This definition is only applicable if there is a carrier. We will use the following more general definition for the relative bandwidth η,

$$\eta = (f_H - f_L)/(f_H + f_L) \tag{1}$$

where f_H is the highest and f_L the lowest frequency of interest. Typical radio signals used for communications or radar have relative bandwidths in the order of $\eta = 0.01$ or less, an amplifier specified from $f_L = 0.1$ GHz to $f_H = 2$ GHz yields $\eta = (2 - 0.1)/(2 + 0.1) = 0.9$, and an attenuator specified from dc to 2 GHz yields $\eta = 1$, which is the largest value permitted by the definition. The distortion-free lines discussed in the preceding section have $\eta = 1$. Many commercially available components, such as hybrid couplers or frequency-independent antennas, have relative bandwidths either close to 1 or at least much larger than the typical relative bandwidths of signals. Generally speaking, only circuits and structures designed to resonate with (almost) sinusoidal signals have a small relative bandwidth, since the phenomenon of resonance disappears with increasing relative bandwidth.

For an explanation of this statement refer to Fig. 1.4-1, which shows on top a parallel resonant circuit. The impedance Z presented by it to a sinusoidal current is given by the formula

$$Z = [1/R + j(\omega C - 1/\omega L)]^{-1} \tag{2}$$

which is usually written in a normalized form:

$$Z/Z_0 = [1 + jQ(\Omega - 1/\Omega)]^{-1} \tag{3}$$

$$Z_0 = R, \qquad \omega_0 = (LC)^{-1/2}, \qquad \Omega = \omega/\omega_0, \qquad Q = \omega_0 RC$$

The higher and lower half-power frequencies Ω_H and Ω_L follow from the condition $Q(\Omega - 1/\Omega) = 1$:

$$\Omega_H = 1/2Q + (1 + 1/4Q^2)^{1/2}, \qquad \Omega_L = -1/2Q + (1 + 1/4Q^2)^{1/2} \tag{4}$$

The difference and the sum of Ω_H and Ω_L define the half-power bandwidth and the sum frequency:

$$\Omega_H - \Omega_L = 1/Q, \qquad \Omega_H + \Omega_L = 2(1 + 1/4Q^2)^{1/2} \tag{5}$$

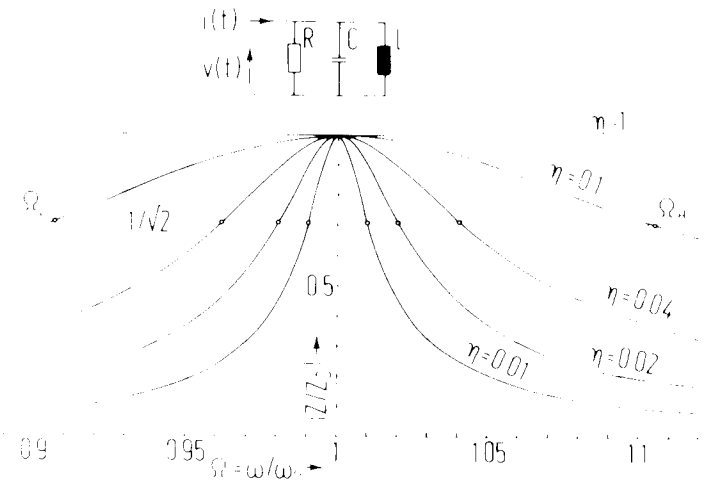

FIG. 1.4-1. The magnitude of the relative impedance Z/Z_0 of a parallel resonant circuit for various values of the relative bandwidth η as function of the normalized frequency $\Omega = \omega/\omega_0$.

The relative half-power bandwidth η is connected with the quality factor Q through Eq. (5):

$$\eta = \frac{\Omega_H - \Omega_L}{\Omega_H + \Omega_L} = \frac{\omega_H - \omega_L}{\omega_H + \omega_L} = \frac{f_H - f_L}{f_H + f_L} = \frac{1}{2Q(1 + 1/4Q^2)^{1/2}} \tag{6}$$

The magnitude of the normalized impedance and the half-power bandwidth can now be written in terms of η:

$$\left|\frac{Z}{Z_0}\right| = \left[1 + \left(\Omega - \frac{1}{\Omega}\right)^2 \frac{1 - \eta^2}{4\eta^2}\right]^{1/2} \tag{7}$$

$$\Omega_H - \Omega_L = 2\eta(1 - \eta^2)^{-1/2} \tag{8}$$

The normalized impedance $|Z/Z_0|$ and the bandwidths $\Omega_H - \Omega_L$ are shown in Fig. 1.4-1 for various values of the relative bandwidth η. One may readily see that η must be of the order of 0.01 or less to obtain an appreciable resonance effect. Similar results hold for other circuits or structures resonating with sinusoidal functions, in particular for the resonant dipole. This behavior is the main reason why our usual technology for radio transmission requires small relative bandwidths.[1]

The resonant circuits for sinusoidal pulses, Legendre pulses, and Bessel functions in Figs. 1.2-3, 1.2-7, and 1.2-10 do not have a small relative (fre-

[1] Radars have been built with a relative bandwidth η as high as 0.05, using the conventional technology based on sinusoidal functions. It is clear from Fig. 1.4-1 that the selectivity and noise rejection of such equipment cannot be good. We have here a classical example of how the lack of an adequate theory leads to enormous investments that eventually have to be written off.

quency) bandwidth since the Fourier transform of the signals for, e.g., $i = 1$ in Fig. 1.2-2, $j = 1$ in Fig. 1.2-6, and $n = 0$ in Fig. 1.2-9 contains important frequency components close to zero. However, this is so only because the usual Fourier transform and the usual concept of bandwidth are based on the periodic sinusoidal functions. One can generalize the Fourier transform and the concept of bandwidth by basing them on other systems of functions (Harmuth, 1969, 1972). The concept of relative bandwidth then becomes generalized too, and all resonant filters for a specific system of functions have a small relative bandwidth in terms of that system. This is not so for the resonant filter of Fig. 1.3-1, which resonates with the large class of periodic functions with period T.

The use of signals with a small relative bandwidth is obviously a restriction imposed on us by the current technology, but it was not felt as a restriction until very recently. Consider, e.g., a radio signal in the AM band from 535 to 1605 kHz. It has necessarily a small relative bandwidth, since our ears respond only to frequencies below about 16 kHz. A problem would have arisen if we had tried to transmit music with a radio carrier having a frequency of, e.g., 20 kHz, but there was not much demand for such a service.

The signals we want to transmit have usually a relative bandwidth close to 1. For instance, audible signals use the band from about 20 Hz to about 16 kHz. Their relative bandwidth is thus $\eta = (16{,}000 - 20)/(16{,}000 + 20) = 0.9975$. Even a low-quality telephone channel transmitting only the band 300 Hz $\leqq f \leqq$ 3000 Hz has a relative bandwidth $\eta = 2700/3300 = 0.82$. Television signals start at dc and thus have a relative bandwidth equal to 1; the same holds true for rectangular pulses used for telegraphy, teletype or data transmission.

The transformation of the large relative bandwidth of these signals into a small relative bandwidth is typically done by means of amplitude modulation of a sinusoidal carrier. Consider the rectangular pulse on top of Fig. 1.2-2 with the duration T. Most of its energy is in the band from $f_L = 0$ to $f_H = 1/T = \Delta f$. Double sideband amplitude modulation of a sinusoidal carrier with frequency f_c shifts this energy to the band from $f_L = f_c - \Delta f$ to $f_H = f_c + \Delta f$. The relative bandwidth

$$\eta = \frac{(f_c + \Delta f) - (f_c - \Delta f)}{(f_c + \Delta f) + (f_c - \Delta f)} = \frac{\Delta f}{f_c} = \frac{1}{f_c T} \tag{9}$$

can in principle be made as small as one wants by choosing f_c sufficiently large compared with $1/T$. Let f_c equal $1/T$, $2/T$, $3/T$, ... The resulting sinusoidal carriers modulated by the rectangular pulse are the sinusoidal pulses with $i = 1, 2, 3, \ldots$ in Fig. 1.2-2. As i increases, the sinusoidal pulses look more and more like periodic sinusoidal functions. Hence, a

small relative bandwidth means that a signal looks very similar to a periodic sinusoidal function.

The process of turning the generally large relative bandwidth of signals into a small relative bandwidth by means of a large carrier frequency f_c will fail, if conditions imposed by nature prevent us from choosing f_c as large as we want. We will investigate when this happens in the following section.[1]

The use of large relative bandwidths runs counter to current regulations based on CCIR recommendations. Tables 1.4-1 and 1.4-2 show the relative bandwidths of broadcasting and radio-location channels in the United States. Some of the broadcasting channels, and particularly the combination of adjacent channels, have fairly large relative bandwidths, but only two radio-location channels have a relative bandwidth slightly larger than 0.1. This seems to imply insurmountable regulatory obstacles. However, this is not so. CCIR Question 1A/29 asks specifically for a study of "what are technical criteria for sharing bandwidth expansion with conventional modulation systems and what are appropriate techniques and design factors conducive to improving the efficiency of spectrum utilization in

TABLE 1.4-1
RELATIVE BANDWIDTHS OF BROADCASTING BANDS ACCORDING TO THE UNITED STATES FREQUENCY ALLOCATIONS[a]

Service	Occupied band (MHz)	Relative bandwidth			
AM radio	0.535–1.605	0.5			
TV channels 2–4	54–72	0.14	0.24		
TV channels 5–6	76–88	0.07		0.17	0.33
FM radio	88–108	0.10			
TV channels 7–13	174–216	0.11			
TV channels 14–20	470–512	0.04	0.26		
TV channels 21–69	512–806	0.22			

[a] The relative bandwidth is shown for the individual bands as well as for certain combinations, e.g., 0.17 for TV channels 5–6 plus FM radio.

[1] One school of thought represented, e.g., by W. H. Kummer of Hughes Aircraft Co., holds that only the absolute bandwidth but not the relative bandwidth is significant, since the transmittable information is proportionate to the absolute bandwidth. This is correct within the realm of pure theory, where the properties of the transmission medium for the electromagnetic waves can be defined by the investigator to produce the desired result. The relative bandwidth is only significant for practical applications, where the features of the transmission medium are imposed on us by nature. The following Section 1.5 shows this for radars operating in the atmosphere. Similar restrictions hold for ice, fresh water, sand, clay, and other media encountered by the waves of the into-the-ground radar that will be discussed in Section 1.6. The restrictions imposed by saltwater will be discussed in Section 1.6 and in more detail in Chapter 7.

TABLE 1.4-2

RELATIVE BANDWIDTHS OF RADIO-LOCATION CHANNELS ACCORDING TO THE UNITED STATES FREQUENCY ALLOCATION

Band (MHz)	216–225	420–450	902–928			
Relative bandwidth	0.020	0.034	0.014			
Band (GHz)	1.215–1.4	2.3–2.5	2.7–3.7	5.25–5.925	8.5–10.55	13.4–14
Relative bandwidth	0.071	0.042	0.16	0.060	0.11	0.022
Band (GHz)	15.7–17.7	33.4–36	48–50	71–76	165–170	240–250
Relative bandwidth	0.060	0.037	0.020	0.034	0.015	0.020

shared frequency bands'' (CCIR, 1975). A number of reports have been written in response to this question (e.g., U.S.A., 1977a–c), and the CCIR recommendations will eventually be changed to permit frequency sharing services, which will allow the general use of large relative bandwidths. At the present time, the use of large relative bandwidths is quite legitimate if the radiation is sufficiently localized. This condition is met by the commercially available into-the-ground radars. Furthermore, the spreading of power over frequency bands with a width up to 10 GHz reduces the power per unit bandwidth—or the power density—to such low levels that the radiation cannot be detected by conventional monitoring equipment using a small relative bandwidth. There is, of course, no objection to radiation that is not detectable by conventional equipment.

1.5 Attenuation of Waves, Noise, and Distortions

Figure 1.5-1 shows the round-trip attenuation of sinusoidal waves due to rain or fog as function of the frequency for the frequency range from 1

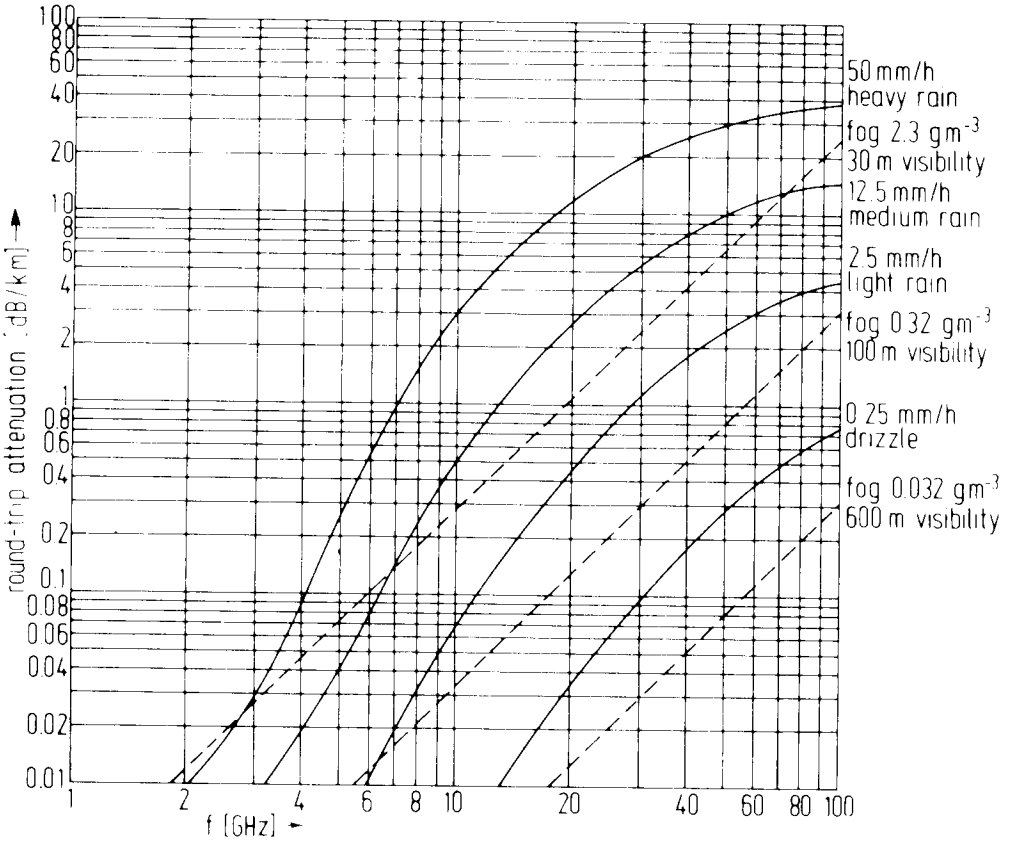

FIG. 1.5-1. Round-trip attenuation of a radar signal in decibels per kilometer distance to the target for the frequency range from 1 to 100 GHz. Curves of this type seem to have been obtained first by J. W. Ryde and D. Ryde (Ridenour, 1947, p. 61). They may be found in various presentations in many publications. The curves for fog were taken from p. 472 of the book by Barton (1964). The curves for rain were plotted from data computed by Setzer (1970); they differ slightly at low frequencies from corresponding curves published by Barton (1964, p. 472).

to 100 GHz. This illustration is specifically intended for radar. Round trip means that the waves travel the distance $2D$ if the target has the distance D from a radar.[1] The plot is not extended below 1 GHz since the attenuation becomes negligible there.

Figure 1.5-2 shows the noise temperature in the range from 100 MHz to 100 GHz for a radar with the elevation angles 0°, 1°, ..., 90°, where 0° means that the radar looks toward the horizon, and 90° means that it looks vertically up.

Figure 1.5-3 shows the round-trip attenuation of sinusoidal waves due to molecular absorption at sea level as a function of the frequency for the frequency range $0 \leqq f \leqq 130$ GHz. There is essentially no attenuation below about 10 GHz. Regions with low attenuation occur around 35 and 94 GHz. Peaks occur at 22.2, 60, and 118 GHz due to resonances of the H_2O and O_2 molecules.[2]

Let us use these three illustrations to determine when a signal should not be shifted to a higher frequency band in order to achieve a small relative bandwidth. Consider a radar using pulses with a duration of $T = 1\ \mu$s, which yield a range resolution of about 150 m. The pulses occupy the approximate band $0 \leqq f \leqq 1$ MHz. In order to turn the absolute bandwidth $\Delta f = 1$ MHz in the baseband into a relative bandwidth $\eta = 0.01$ we need a carrier frequency

$$f_c = 1/\eta T = \Delta f/\eta = 100 \text{ MHz}$$

---

[1] This illustration provides a good example of how the concept of small relative bandwidth enters our thinking. The attenuation of a pure sinusoidal wave between two points with distance D or $2D$ is a well-defined concept. There are no distortions, whereas the change of attenuation with frequency implies distortions for all other signals. The usual conclusion is that one should use signals with small relative bandwidth in order to keep the distortions low. It is rarely recognized that the price for no distortions is no range resolution, and the price for low distortions is poor range resolution.

[2] Data on molecular absorption may be found in many references (Altshuter *et al.* 1968; Falcone and Abreu, 1979; Moore 1979; U.S.A. 1977a,c). Some of these references also contain plots for rain and fog. The plots in Fig. 1.5-1 for fog correspond with data by Chen (1975) for *coastal fog;* the attenuation for *inland fog* is about half as large. Data for frequencies higher than used here are given by Zhevakin and Naumov (1967) and by Deirmendjian (1975). One should observe that the plots of Figs. 1.5-1–1.5-3 contain many simplifying assumptions that are needed to permit us to draw a few simple plots for "typical" conditions. For instance, the attenuation for rain in Fig. 1.5-1 depends not only on the rate of rainfall, but on the size of the drops, which may vary for the same rate of rainfall. The plots of Fig. 1.5-2 also depend on the reception angle of the antenna, the concentration of water vapor in the air, etc. The plot of Fig. 1.5-3 depends again on the concentration of water vapor in the air, on the air pressure, etc. The reader interested in such finer details is referred to the cited literature. The huge differences between attenuation and distortion in the bands 0.5 GHz $\leqq f \leqq$ 10.5 GHz and 89 GHz $\leqq f \leqq$ 99 GHz discussed here obviate the need to discuss these finer details.

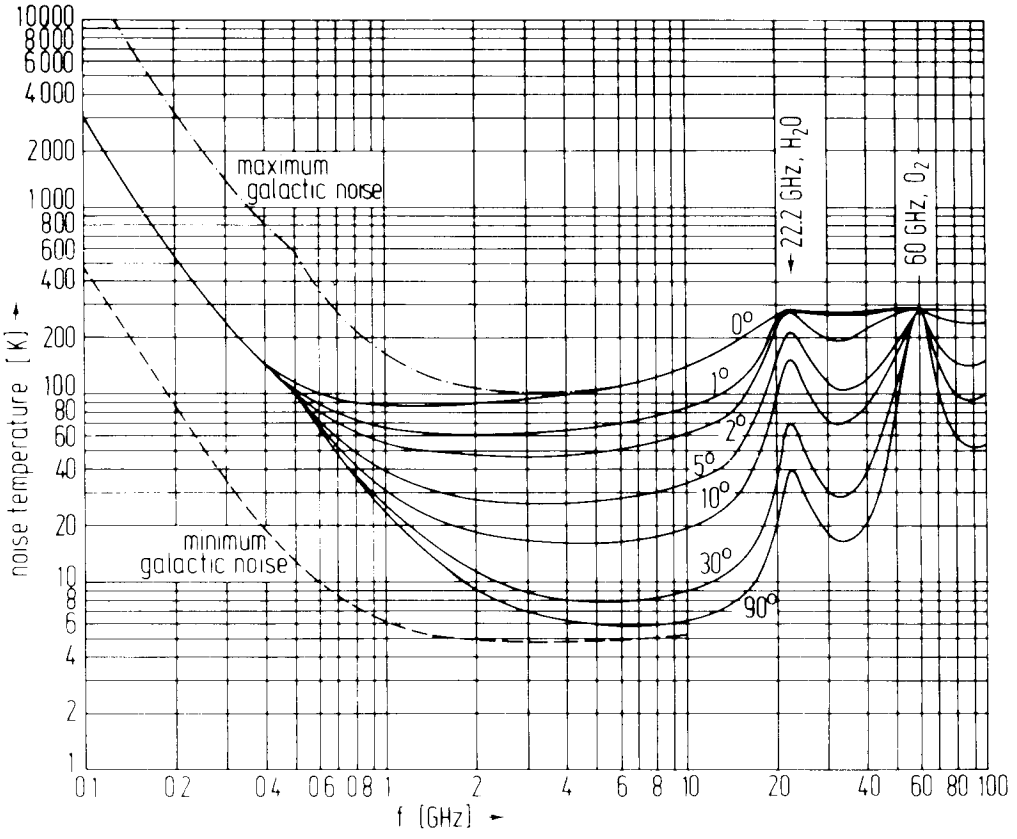

FIG. 1.5-2. Noise temperature as function of the frequency for the elevation angles 0°, 1°, ..., 90° of a radar. These plots are due to L. V. Blake of the Naval Research Laboratory. They may be found in a number of reports, papers, and books, e.g., in *Radar Handbook* (Blake, 1970).

according to Eq. (1.4-9). At 100 MHz, the noise temperature is still very high according to Fig. 1.5-2. One would much rather operate at a higher frequency, preferably in the range from 1 to 10 GHz. Hence, the requirement for a small relative bandwidth is overshadowed by the desire to operate at a higher frequency, where the noise temperature is lower. Furthermore, the radar dish assumes a more manageable size when one advances from 100 MHz to a higher frequency.

Let us next consider a radar that operates with pulses having a duration between 1 and 0.1 ns, which implies a range resolution between 15 and 1.5 cm. The baseband bandwidth Δf of the pulses is between 1 and 10 GHz. A relative bandwidth $\Delta f/f_c = 0.01$ calls for a carrier frequency between 100 and 1000 GHz. Figure 1.5-2 shows that the noise temperature at such high frequencies is well above the value in the band from 1 to 10 GHz. The real objection, however, does not come from the noise temperature but from the attenuation due to moisture and molecular absorption.

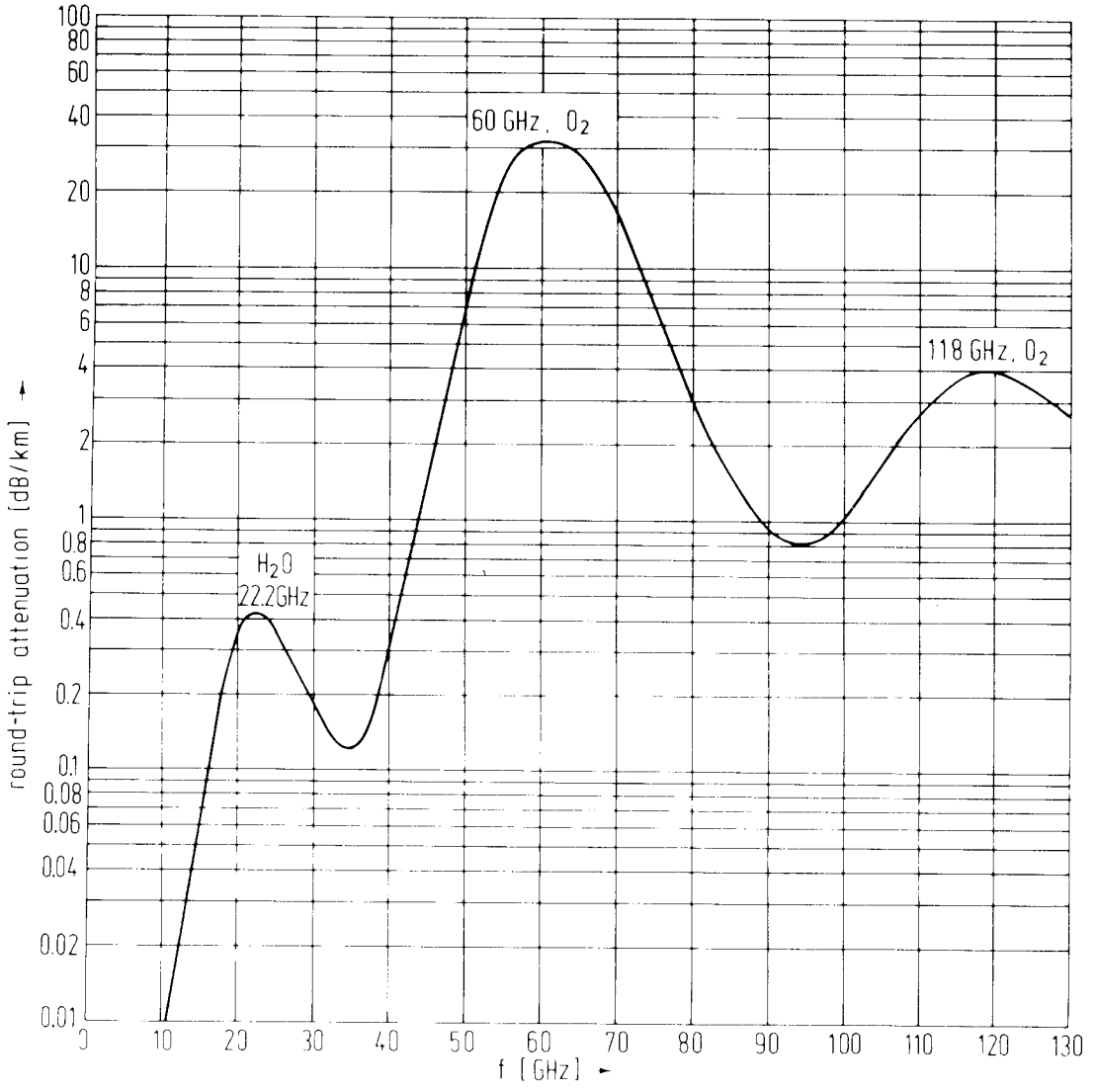

FIG. 1.5-3. Round-trip attenuation of sinusoidal waves due to molecular absorption at sea level. This plot is based primarily on data published by Zhevakin and Naumov (1967) and Blake (1970): The water vapor content of the air is assumed to be 7.5 g/m³.

Let us turn to Fig. 1.5-1 and consider a radar[1] using an absolute bandwidth of 10 GHz at the location 89 GHz $< f <$ 99 GHz. The relative bandwidth $\eta = 10/(89 + 99) \doteq 0.05$ is so large that there is only a very moderate resonance effect according to Fig. 1.4-1, and the noise rejection

[1] This reference system was originally advocated by D. K. Barton of Raytheon Co., a leading proponent of conventional techniques, as an example of what could be achieved by staying with small relative bandwidths. It now provides the best arguments for the use of large relative bandwidths. The pulse duration of 0.1 ns and the distance of 10 km are used because of the practical importance of these values for the low-elevation angle radar and the ICBM radar. More details on applications of radar operating at 35 or 94 GHz may be found in a publication by Wiltse (1979), and the many references given there.

by the filter is thus rather poor. The attenuation in fog of 30 m visibility for the signal is about 250 dB for a distance of 10 km to the target (Fig. 1.5-1 shows about 25 dB/km at $f = 100$ GHz). If, on the other hand, the signal with an absolute bandwidth of 10 GHz is at the location 0.5 GHz $<$ $f <$ 10.5 GHz, one obtains a relative bandwidth $\eta = 10/(0.5 + 10.5) \doteq 0.9$. The attenuation at the upper band edge for a distance of 10 km is now 3 dB according to Fig. 1.5-1, and negligible at the lower band edge. The net gain due to the increase of the relative bandwidth is thus about[1] 247 dB.

Let us turn to Fig. 1.5-3. There is essentially no attenuation for the signal in the band 0.5 GHz $\leqq f \leqq$ 10.5 GHz, but the attenuation for the signal in the band 89 GHz $\leqq f \leqq$ 99 GHz equals more than 0.8 dB/km, or 8 dB for a distance of 10 km. Combined with the 247 dB due to fog we obtain an advantage for the large-relative-bandwidth radar over the conventional radar pushed to its limits of about 255 dB in the presence of heavy fog, and about 8 dB if the weather is fair.

It is worth contemplating the gain of 255 dB, since it means a factor of 3×10^{25}. The reader will probably not be able to come up with another example of a theoretical insight providing anywhere near such an improvement—be it in power or anything else—over an existing equipment or method.[2] To call this improvement astronomical would be an understatement, since we would have to give the distance to the farthest known celestial objects in meters or the known age of the universe in microseconds to obtain a number of the order of 10^{25}.

We return again to the question of noise. According to Fig. 1.5-2 the lowest noise temperature is in the region from about 1 to about 10 GHz, while the noise temperature in the band 89 GHz $< f <$ 99 GHz is significantly higher. However, the question is sometimes raised whether signals with large relative bandwidth are equally affected by noise as signals with small relative bandwidth. This question was answered by Kotel'nikov (1947) some 30 years ago. For additive thermal noise the time variation of the signal is quite unimportant, only the energy of the signal affects the error probability,[3] and the time variation is generally the only difference between any radio signals having the same polarization and power.

---

[1] Similar considerations apply, if the attenuation is due to chaff rather than water droplets. A pulse with a duration of 0.1 μs should penetrate chaff about like a pulse with a duration of 0.1 ns penetrates rain and fog, if no carrier is used.

[2] A frequency of 100 GHz implies a wavelength of 3 mm. The *millimeter-wave radar* is sometimes advocated for all-weather use (e.g., Whalen 1979). Millimeter waves are substantially less attenuated by fog than infrared or visible waves, but they are slightly more attenuated by rain and substantially more by molecular resonances.

[3] A derivation of this result using the same terms as this book was previously published (Harmuth 1972, pp. 292–299).

Another question often raised is that of signal distortions. In vacuum, the propagation of an electromagnetic wave is described by the wave equation, which implies distortion-free transmission. In the atmosphere, for transmission above the shortwave region, distortions may be caused by rain, fog, and molecular resonances. It was just discussed that a signal in the band 0.5 GHz $< f <$ 10.5 GHz suffers an attenuation of 3 dB at the upper band edge in heavy fog, whereas there is essentially no attenuation at the lower band edge. This implies a distortion that may be corrected, if one feels a need to do so, by an equalizer having 3 dB more attenuation at 0.5 GHz than at 10.5 GHz.

The signal in the band 89 GHz $\leqq f \leqq$ 99 GHz has an attenuation of about 20 dB/km at the lower band edge and of about 25 dB/km at the upper band edge according to Fig. 1.5-1. The difference equals 5 dB/km, or 50 dB for a distance of 10 km. Hence, the attenuation distortion increases from 3 to 50 dB if we replace the signal with large relative bandwidth with the one having a small relative bandwidth. This is just the opposite of what most communication engineers would expect.

The question of distortions due to molecular resonances can be answered with the help of Fig. 1.5-3. There is essentially no absorption and thus no distortion in the band 0.5 GHz $\leqq f \leqq$ 10.5 GHz. In the band 89 GHz $\leqq f \leqq$ 99 GHz the round-trip attenuation varies by about 0.2 dB/km, or 2 dB for a distance of 10 km. Hence, molecular resonances cause insignificant distortions in either case. The situation would be completely different for a radar operating in the band 40 GHz $\leqq f \leqq$ 50 GHz, where the attenuation varies by about 7 dB/km.

From all this it is clear that the region from about 0.5 GHz to somewhat above 10 GHz is ideal for an all-weather radar, and that signals with a bandwidth of about 10 GHz can be used even under the worst weather conditions. In fair weather, a large-relative-bandwidth radar can use a bandwidth close to 45 GHz, and have an attenuation below that of a narrow-band radar operating at 94 GHz according to Fig. 1.5-3.[1]

---

[1] The introduction of nonsinusoidal waves to radar aroused many objections from engineers and scientists involved in the manufacture and development of conventional radars. The scientific reasons for these objections have two sources: (a) It is not understood that Maxwell's equations do not distinguish sinusoidal waves, although solutions obtained with Bernoulli's product method do so, as discussed at the end of Section 1.3. (b) It is not recognized that most of the existing radar and radio equipment works best or even only with sinusoidal waves because it was designed to do so. In order to provide a public forum for a discussion of objections to nonsinusoidal waves, a panel discussion on "Radio Signals with Large Relative Bandwidth" was organized on 7 October 1980 at the IEEE International Symposium on Electromagnetic Compatibility in Baltimore, Maryland. Calls for opponents were published in *Microwave J.* **23** (No. 1, Jan. 1980, p. 10) and *IEEE Trans. Electromagn. Compat.* **EMC-22** (No. 1, Feb. 1980, p. 86). Many personal invitations were sent to known

1.6 When to Use Nonsinusoidal Waves

The frequency band from about 1 to about 10 GHz with a bandwidth of essentially 10 GHz has just been identified to be ideally suited for signals with large relative bandwidth, or *nonsinusoidal signals* for short. The typical application for this band is the line-of-sight, high-resolution, all-weather radar.

The conventional line-of-sight, all-weather radar cannot use signals with a bandwidth wider than about 100 MHz in the baseband if the highest frequency after modulation is restricted to about 10 GHz. The possible bandwidth of the radar with large relative bandwidth is about 100 times as large, which implies an increase in information acquisition by a factor of about 100. This gain consists inherently in better time resolution—the width of the used pulses can be decreased from 10 to 0.1 ns—but time resolution can be translated into range resolution, Doppler resolution, angular resolution, target signature, etc.

Radar is the major field for the use of very large absolute bandwidths. A lesser field is spread spectrum communications. Nonsinusoidal waves permit us to spread signals over a bandwidth of 10 GHz, stay below the resonance of the H_2O molecule, and avoid the attenuation by rain and fog.

To find other useful ranges in the frequency spectrum we must find instances where an increase of the frequency of a carrier is essentially impossible. The shortwave band is an example. Using a typical relative bandwidth of $\eta = 0.01$ and a carrier frequency $f_c = 15$ MHz in the middle of the shortwave band, we obtain from Eq. (1.4-9) the baseband bandwidth $\Delta f = \eta f_c = 150$ kHz. Since the shortwave band inherently permits the transmission of about 3 MHz wide signals before the frequency dependence of propagation becomes a problem, we use less than 10% of the bandwidth allowed by nature if we insist on a small relative bandwidth. Using techniques for large relative bandwidths, we may increase the absolute bandwidth of signals by about one order of magnitude compared with the typical case $\eta = 0.01$. The major use for this potential gain is for over-the-horizon radar, where the limited resolution has been a

---

opponents in industry and government, but none were accepted. The discussion panel, chaired by the author, consisted thus only of proponents: C. Bertram of Bertram Technology Inc., Merrimac, New Hampshire; J. C. Cook of Teledyne-Geotech, Dallas, Texas; A. Dean of U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire; Fan Changxin of Northwest Telecommunication Engineering Institute, Xi'an, China; H. Lueg of Technische Hochschule Aachen, West Germany; J. V. Rosetta of Geophysical Survey Systems Inc., Hudson, New Hampshire. Two more proponents had planned to participate but could not come: J. R. Rossiter of Memorial University of Newfoundland, Canada; M. Zecha of Deutsche Akademie der Wissenschaften, Zentralinstitut für Kybernetik, Berlin, East Germany. A radar produced by Geophysical Survey Systems, Inc. was demonstrated at this symposium by J. V. Rosetta.

problem for many years. Note that the potential gain in absolute bandwidth for the over-the-horizon radar is only one order of magnitude, whereas it was two orders of magnitude for the line-of-sight radar.

The attenuation of an electromagnetic wave in seawater is given by the formula

$$\alpha(f) = 0.0345\sqrt{f} \quad [\text{dB/m}] \tag{1}$$

where f is measured in hertz. For a penetration depth of 300 m, which is typically required for radio communication with deeply submerged submarines, one cannot use frequencies higher than some tens of hertz; e.g., for $f = 100$ Hz one obtains from Eq. (1) already an attenuation of 90 dB for a depth of 300 m. With the highest possible frequency so severely limited, the restriction of the relative bandwidth to values of about $\eta = 0.01$ would result in absolute bandwidths for a signal on the order of fractions of a hertz, and a corresponding small transmission rate of information. Hence, radio communications to and from deeply submerged submarines, and radar that penetrates seawater, are among the most intriguing applications of nonsinusoidal signals.

A very similar application is the into-the-ground radar. Inhomogeneities in the ground make the attenuation of electromagnetic waves increase with frequency very rapidly, just like raindrops and fog droplets make the attenuation in air increase very rapidly with frequency according to Fig. 1.5-1. In order to provide an acceptable resolution, such into-the-ground radars must use pulses with a duration of about 1 ns. The penetration depth currently achieved is a few meters in dry ground and at least 30 m in ice (Rossiter and Butt, 1979); the use of a carrier and a small relative bandwidth would result in no useful penetration. Some 60 such into-the-ground radars have been built so far, and they have been used in ten countries.[1] Their typical use is the probing of the ground in the course

[1] The following commercial organizations have developed into-the-ground radars: ENSCO Inc., Springfield, Virginia (J. Fowler, L. Davis); Geophysical Survey Systems Inc., Hudson, New Hampshire (J. Mann, J. V. Rosetta, D. F. Stanfill *et al.*); Morey Research Co., Nashua, New Hampshire (R. Morey); MPB Technologies Inc., Ste. Anne de Bellevue PQ, Canada (J. M. Keelty, S. Y. K. Tam); Teledyne-Geotech, Dallas, Texas (J. C. Cook); Terrestrial Systems Inc., Lexington, Massachusetts (J. Chapman). Academic developers and/or users are: Ohio State University Electronics Laboratory (C. W. Davis III, D. L. Moffatt, L. Peters *et al.*); Centre for Cold Ocean Resources Engineering, Memorial University of Newfoundland, Canada (J. R. Rossiter, K. A. Butt *et al.*); Department of Geophysics and Astronomy, University of British Columbia, Canada (G. K. C. Clarke, B. B. Narod); Institut für Technische Elektronik, Technische Hochschule Aachen, West Germany (H. Lueg). Government agencies that have made major contributions are: U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire (J. Brown, A. Dean, A. Kovacs *et al.*); Geological Survey of Canada, Ottawa (A. P. Annan, J. L. Davis, W. J. Scott); Office of Naval Research, Washington, D.C. (H. Bolezalek).

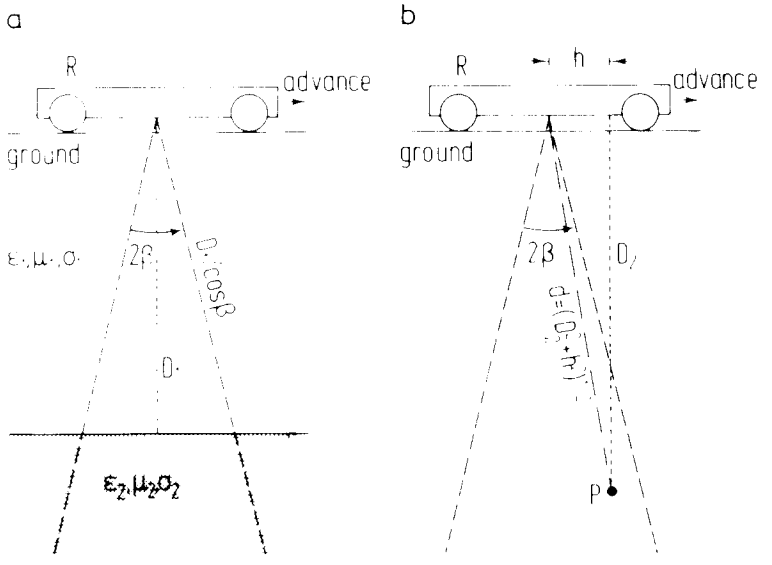

FIG. 1.6-1. Principle of the operation of an into-the-ground radar. (a) Probing of an extensive layer with discontinuity of ϵ, μ, σ; (b) detection of a pipe with small diameter: R, radar; P, pipe.

of excavations, since they will not only detect metallic pipes but also pipes made of concrete, clay or plastic. Other uses have been the detection of cavities in the ground, measuring the thickness of sea ice to assure a safe route for trucks or a landing site for aircrafts on the Arctic Ocean, etc.

The carrier-free into-the-ground radar was first proposed by J. C. Cook[1] (1960); a considerable number of papers on into-the-ground radar with and without carrier have been published since then.[2] For an explanation of the operation of this type of radar, refer to Fig. 1.6-1a. The radar is mounted on a cart that is pulled along the surface of the ground. A pulse of about 1 ns duration is radiated into the ground at intervals of the order of microseconds to milliseconds. A layer at the depth D_1 with a discontinuity of the dielectric constant ϵ, the magnetic permeability μ, or the conductivity σ will reflect the signal. The reflected signal will come from a

[1] By an incredible coincidence, this paper was published in the same year and in the same journal as the author's first papers on nonsinusoidal signals (Harmuth, 1960a,b).

[2] Annan and Davis (1976); Bertram *et al.* (1972); Campbell and Orange (1974); J. C. Cook (1970, 1973); Davis *et al.* (1976); Duckworth (1970); Harrison (1970); Kovacs and Abele (1974); Kovacs and Gow (1975, 1977); Kovacs and Morey (1978, 1979a,b,c); Moffat and Puskar (1976); Morey (1974); Morey and Kovacs (1977); Porcello *et al.* (1974); Rossiter and Butt (1979); Rossiter *et al.* (1979). More publications may be found referenced in the papers listed here. No papers on carrier-free radar or radio signals with large relative bandwidth were published by the more typical journals for radar and radio transmission, in particular, *IEEE Trans. Aerospace and Electronic Systems*, *IEEE Trans. Communications*, *IEEE Trans. Antennas and Propagation*, and *Microwave Journal*.

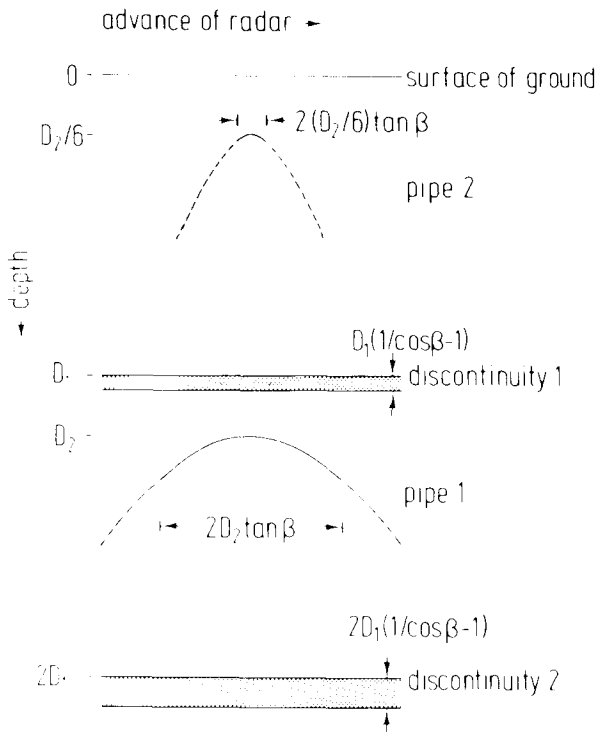

FIG. 1.6-2. Idealized recording of an into-the-ground radar with infinitely short pulses showing the ground profile along the line of advance of the radar. Layers with discontinuity of ϵ, μ, σ are shown at the depths D_1 and $2D_1$; pipes are shown at depths D_2 and $D_2/6$.

minimum distance D_1 and a maximum distance $D_1/\cos\beta$, if the beam angle of the radar equals 2β. Hence, the discontinuous surface is shown in the idealized recording of Fig. 1.6-2 as a band of width $D_1/\cos\beta - D_1$ at the depth D_1; a second discontinuous surface at the depth $2D_1$ produces a band that is twice as wide.

Consider next the detection of a pipe that runs vertically to the paper plane in Fig. 1.6-1 at a depth D_2. If the radar has the horizontal distance h from the pipe, the distance to the pipe will be $d = (D_2^2 + h^2)^{1/2}$. This signature is shown for a pipe 1 at a depth D_2, and for a pipe 2 at a depth $D_2/6$ in Fig. 1.6-2. Due to the beam angle 2β this signature will only be recorded for a maximum horizontal distance $h_{max} = D_2 \tan\beta$ or $h_{max} = (D_2/6)\tan\beta$ in both directions of the pipe. These sections of the pipe signatures are emphasized by solid lines in Fig. 1.6-2.

An actual recording of an into-the-ground radar is shown in Fig. 1.6-3. On top—i.e., at the surface of the ground—is the usual large signal caused by crosstalk from the transmitter to the receiver. It limits the minimum depth that can be probed, which explains why into-the-ground radars are often placed a considerable distance above ground. There are four bands indicating discontinuous surfaces, the deepest one at a depth of about 2.5 m being identified as the water table. The parabola-like signa-

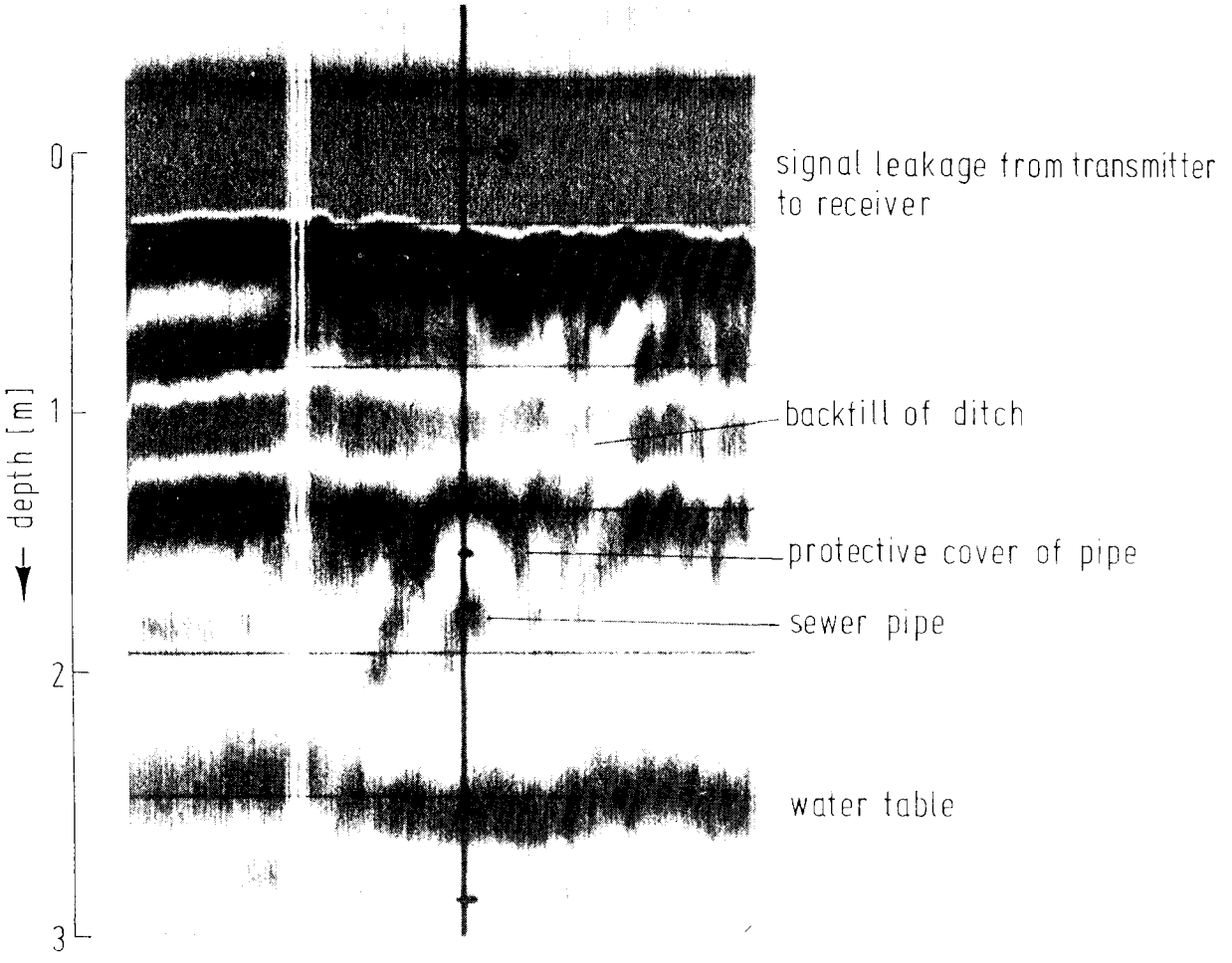

FIG. 1.6-3. Actual recording of an into-the-ground radar. The different signatures of extensive layers and the sewer pipe with its protective cover are conspicuous. (Courtesy J. Chapman of Terrestrial Instruments Inc., Lexington, Massachusetts.)

tures of the sewer pipe and a protective cover over it clearly have the shape shown in Fig. 1.6-2 for pipes. The disturbance above and to the right of the sewer pipe cover signature is caused by construction debris used to back-fill the ditch of the sewer pipe.

Figure 1.6-4 shows a radar profile of a peat deposit. The ground surface and the profile of the clay bottom are clearly visible. The echos just below the surface are caused by buried roots and tree trunks. Two buried trees can be recognized; one of them could be verified visually.

Figure 1.6-5 shows an into-the-ground radar mounted on a cart, as discussed in connection with Fig. 1.6-1, with some processing equipment. For instant displays of ground profiles, as in Figs. 1.6-3 and 1.6-4, one uses graphic recorders. Sometimes the data is stored on magnetic tape and computer processed at a later time.

A radar mounted on a sled and pulled by a tractor that contains the processing and display equipment is shown in Fig. 1.6-6. This radar was used

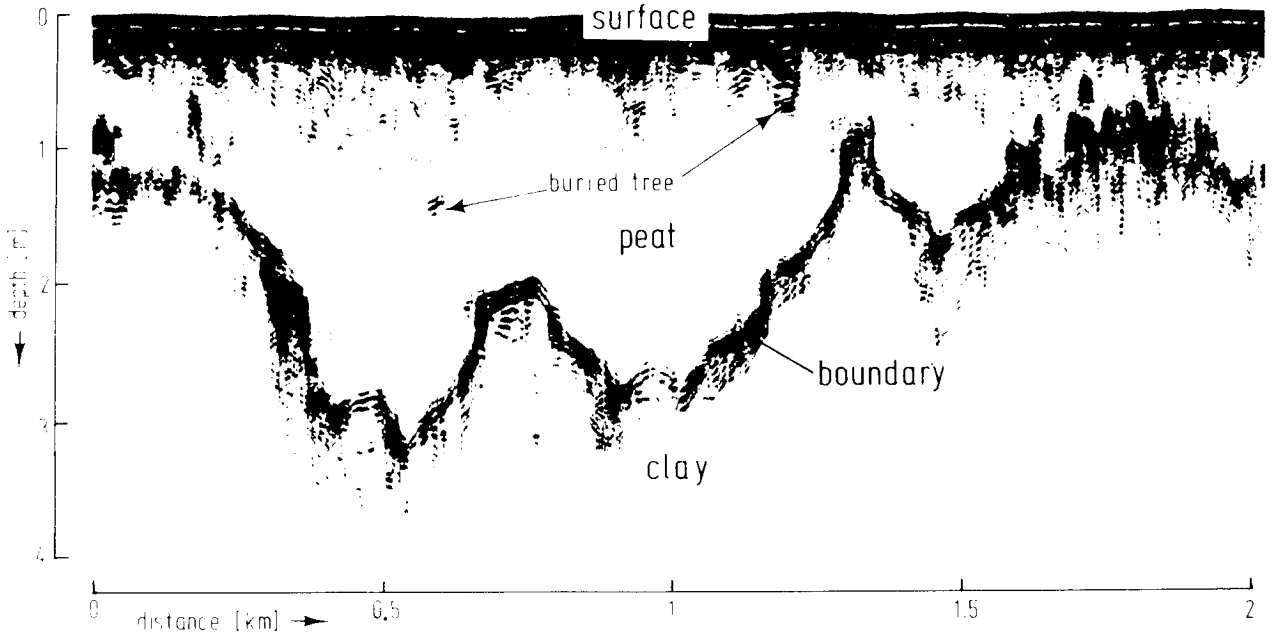

FIG. 1.6-4. Radar profile of a peat deposit. The boundary between the peat and the clay bottom is clearly visible. (Courtesy J. V. Rosetta and D. F. Stanfill of Geophysical Survey Systems Inc., Hudson, New Hampshire.)

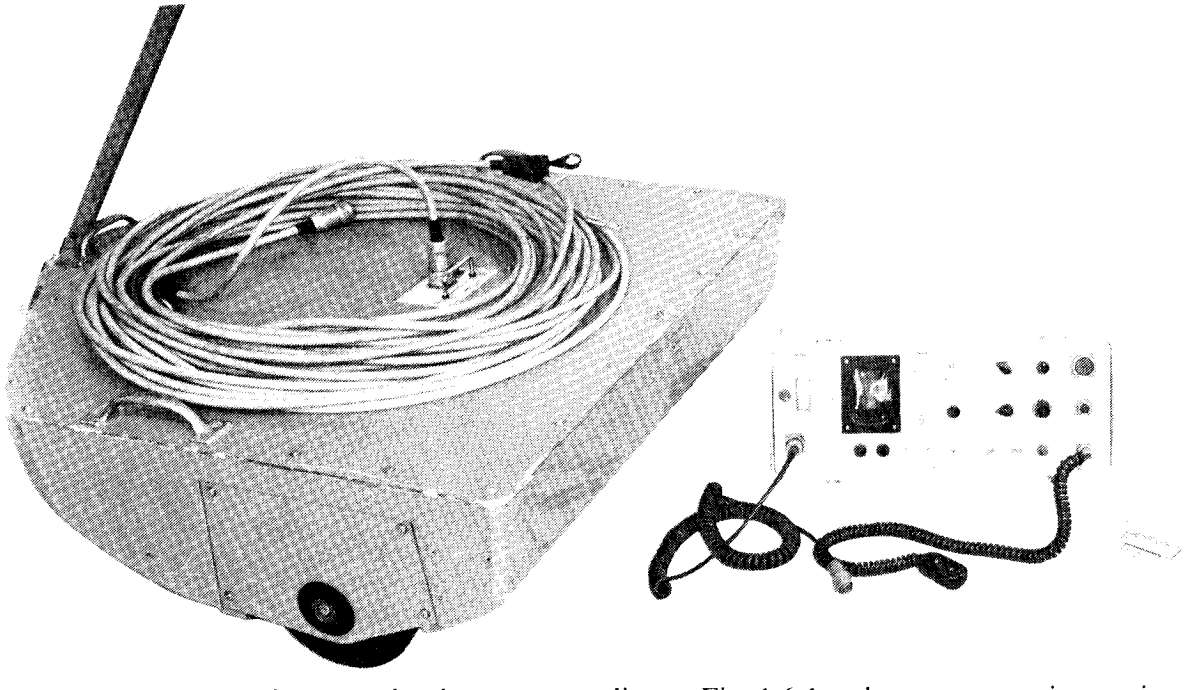

FIG. 1.6-5. Into-the-ground radar cart according to Fig. 1.6-1 and some processing equipment. (Courtesy J. V. Rosetta and D. F. Stanfill of Geophysical Survey Systems Inc., Hudson, New Hampshire.)

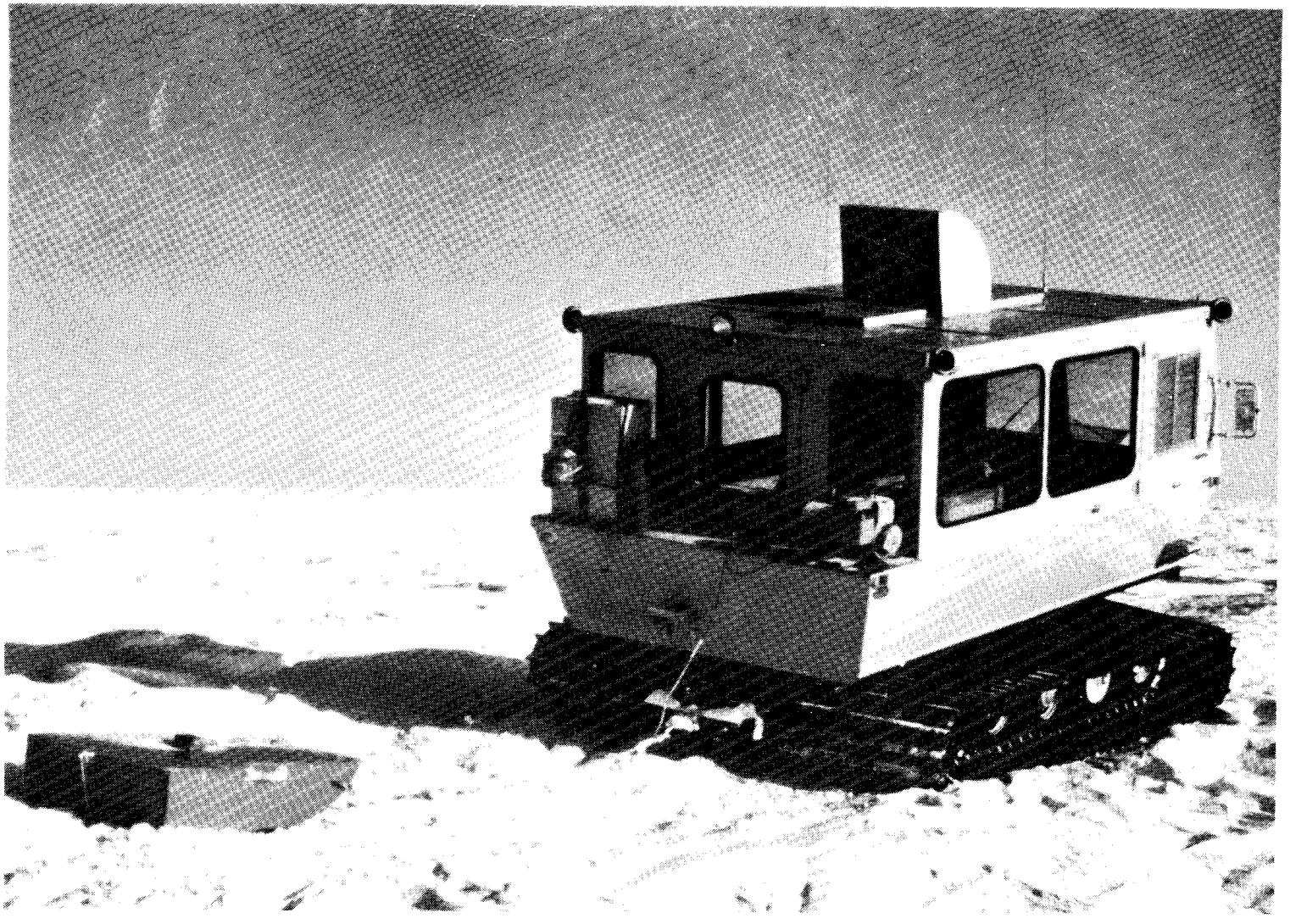

FIG. 1.6-6. Into-the-ground radar mounted on a sled used for surveying a safe truck route on the ice of the Arctic Ocean. (Courtesy J. V. Rosetta and D. F. Stanfill of Geophysical Survey Systems Inc., Hudson, New Hampshire.)

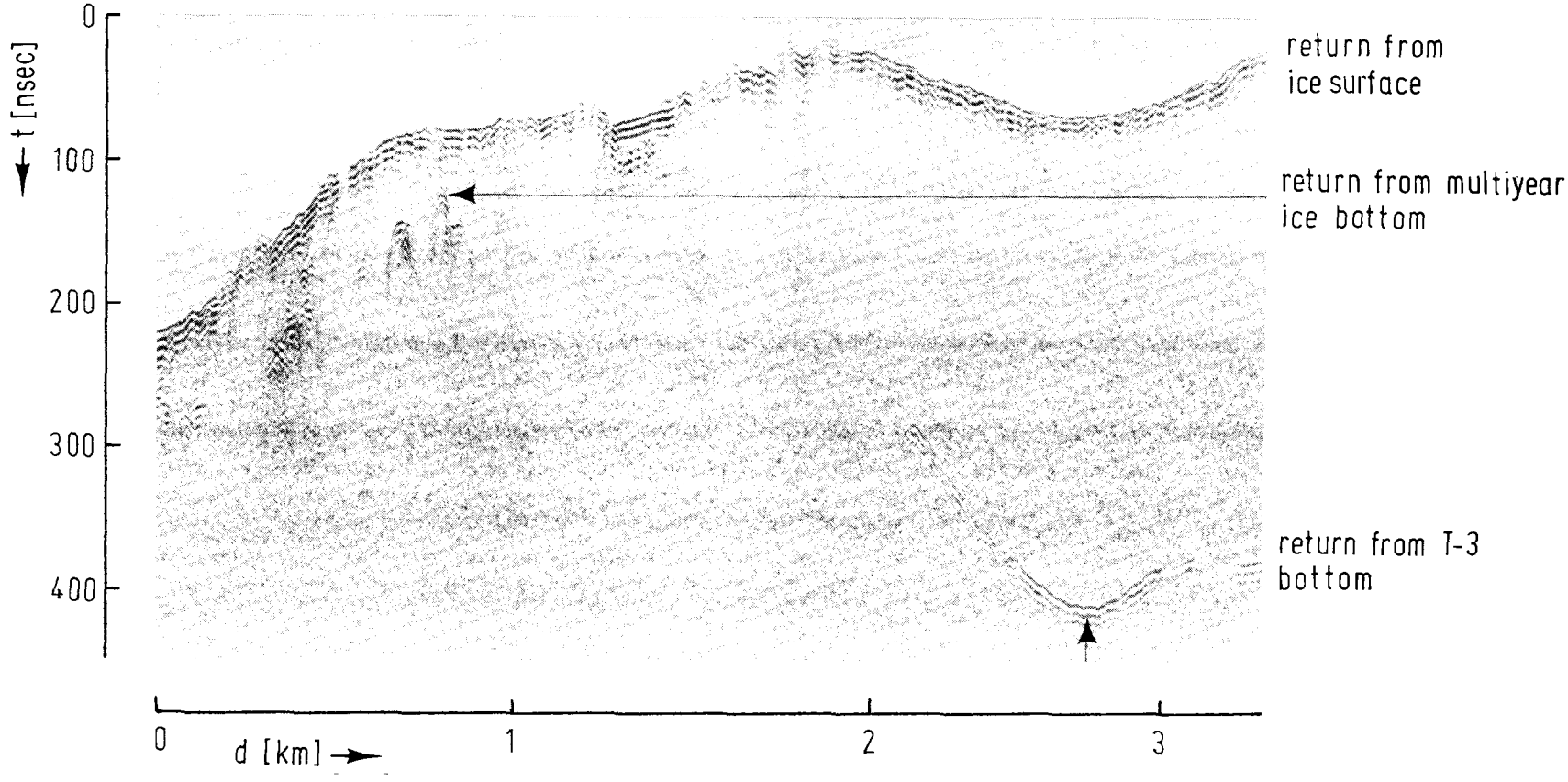

FIG. 1.6-7. Ice profile obtained with a carrier-free radar mounted on an airplane, taken in the Beaufort Sea in March, 1979. The airplane flew at a nominal altitude of 80 m with an airspeed of 70 m/s. The different thickness of single-year and multiyear ice is conspicuous. The return from the bottom of the ice island T-3 at the point indicated by the arrow comes from a depth of 26 m below the ice surface. (Courtesy J. R. Rossiter and K. A. Butt, Centre for Cold Ocean Resources Engineering, Memorial University of Newfoundland, St. John's.)

to probe the thickness of sea ice for the survey of a truck route on the Arctic Ocean.

Ice surveying is a major application of carrier-free radar. It is used either to find thick, safe ice or thin ice that is readily penetrated by ships. Figure 1.6-7 shows an ice profile obtained by an airborne radar in the Beaufort Sea (Rossiter and Butt, 1979). One may distinguish the different thicknesses of first-year ice, multiyear ice, and the ice island[1] T-3. The ice surface appears very uneven, but one must keep in mind that all distances were measured from the aircraft. The uneven surface of the ice indicates that the altitude of the aircraft was not constant along the surveyed line.

Figure 1.6-8 shows the radar with which the ice profile of Fig. 1.6-7 was obtained. This was the first successful use of a carrier-free radar mounted on a fixed-wing aircraft.[2] Helicopter-mounted carrier-free radars have

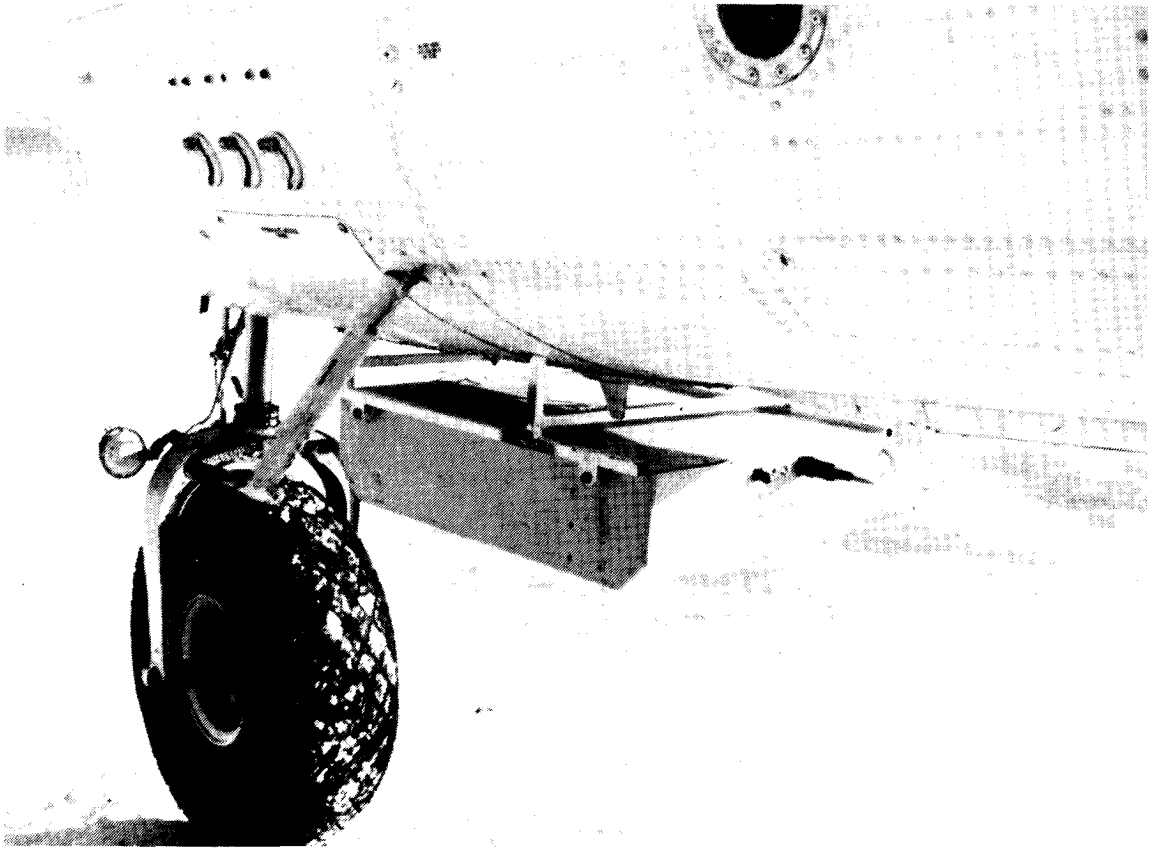

FIG. 1.6-8. Airplane with carrier-free radar used to obtain the ice profile of Fig. 1.6-7. (Courtesy J. R. Rossiter and K. A. Butt, Centre for Cold Ocean Resources Engineering, Memorial University of Newfoundland, St. John's.)

[1] For a description of the ice islands in the Arctic Ocean see Blyth (1953) or Thomas (1965).

[2] The previously mentioned J. C. Cook used an airborne carrier-free radar as early as 1966, but the beamwidth of the antenna was too large to obtain useful results (personal communication; see also Cook, 1970).

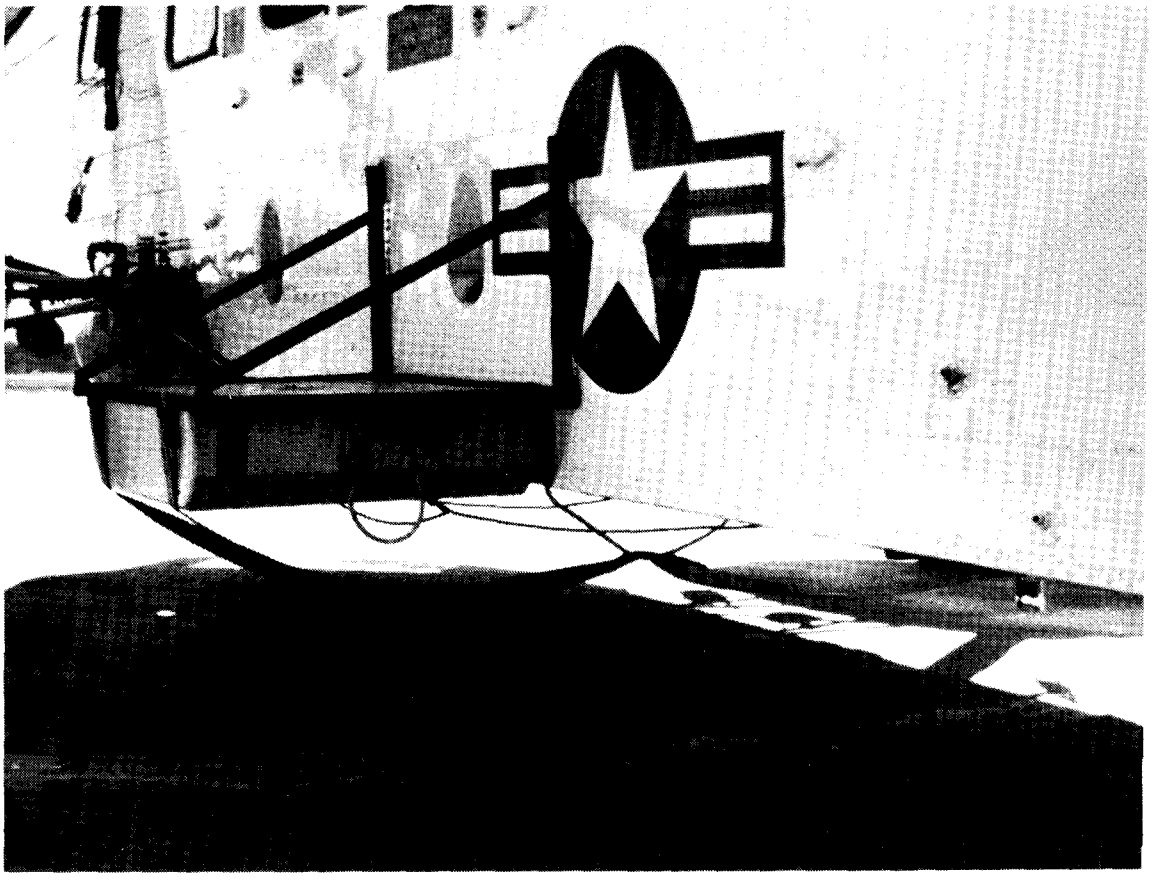

FIG. 1.6-9. Carrier-free radar mounted on a helicopter of the U.S. Coast Guard for ice survey. (Courtesy A. M. Dean, U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire.)

been used since 1976; as far as the author is aware, the first users were A. Dean and A. Kovacs, both of U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire, together with R. Morey of Morey Research Co., Nashua, New Hampshire.[1]

Figure 1.6-9 shows a typical installation of a carrier-free radar on a helicopter of the U.S. Coast Guard used for ice survey of shipping channels. Figure 1.6-10 gives a recording of the St. Marys River below Sault Ste. Marie—between Lake Superior and Lake Huron—obtained with this radar. The sheet ice on the left is clearly distinguishable from the brash

[1] Scientists of the U.S. Navy, Naval Research Laboratory, published a paper in 1979 that tried to prove that there were no useful applications for "carrier-free waveforms" (J. R. Davis *et al.*, 1979). This was three years after the U.S. Coast Guard in cooperation with the U.S. Army had started airborne ice surveys by carrier-free radar, nine years after Geophysical Survey Systems, Inc. had started the commercial production of such radars, and about 15 years after J. C. Cook had built the first experimental into-the-ground radar. Upward of 30 papers reporting experimental results had been published by 1979. We have here a well-recorded example of the time required for a basic new concept to become known and accepted in science.

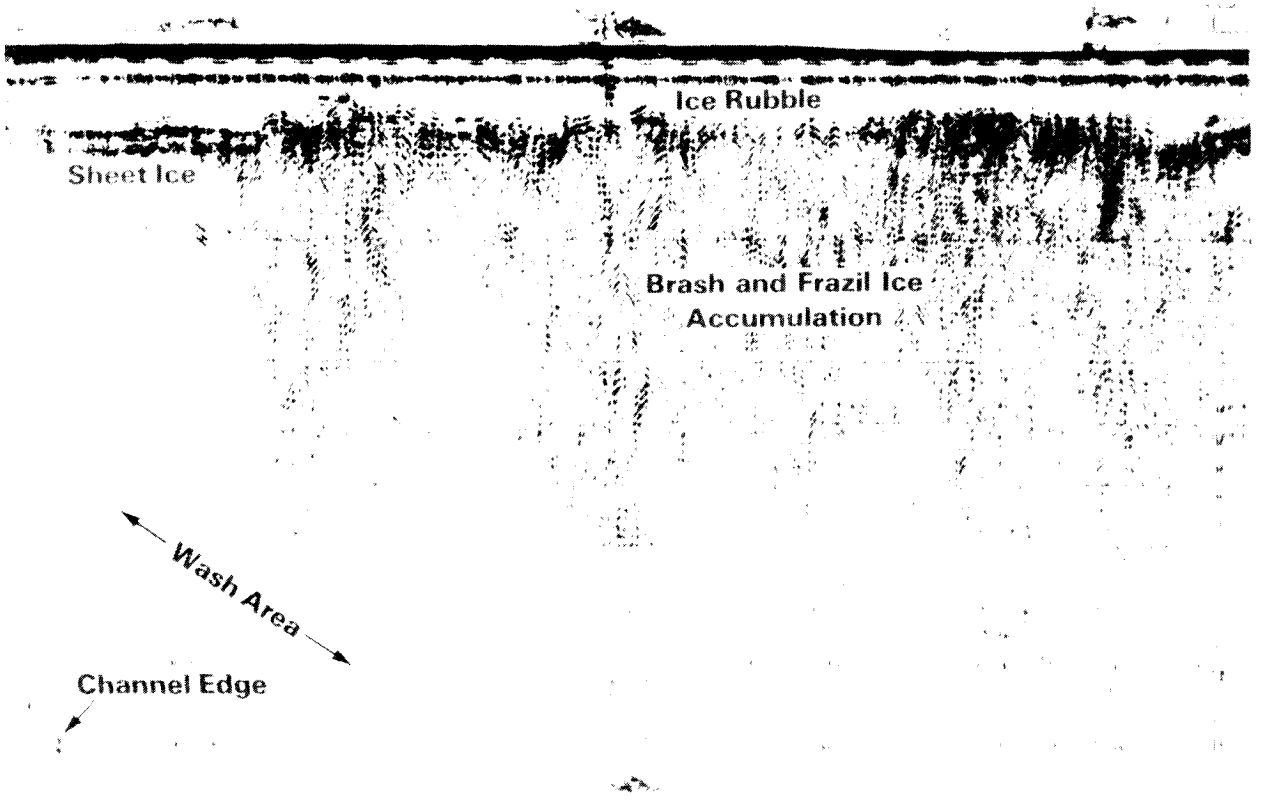

FIG. 1.6-10. Typical recording of the helicopter-borne radar of Fig. 1.6-9, showing a cross section of a shipping channel near Sault Ste. Marie. The ice rubble is about 300 m wide, the brash and frazil ice accumulation some 3–4 m deep. (Courtesy A. M. Dean, U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire.)

and frazil ice accumulation caused by the ship traffic. In the lower left-hand corner one may recognize the channel edge at a depth of about 4 m (the ability of carrier-free radar to penetrate several meters of water, and give a useful resolution, has led to its use for profiling river bottoms from the air to find places that are sufficiently shallow for fording).[1]

The brash and frazil ice accumulation in Fig. 1.6-10 extends to a depth of about 4–5 m. This appears to contradict the depth of 4 m given previously for the channel edge. However, the channel edge in Fig. 1.6-10 is covered by almost 4 m of water—plus the sheet ice—and the electromagnetic wave has a significantly slower propagation velocity in the water than in the ice accumulation to its right. Hence, one must either know or

[1] In the Soviet Union, measurements of the thickness and salinity of sea ice as well as the salinity of seawater by means of airborne radar are carried out by scientists of the Institute of Radio Engineering and Electronics of the Academy of Sciences (Institut Radiotekniki i Elektroniki Ak. Nauk), Karl Marx Prospect 18, 103907 Moscow.

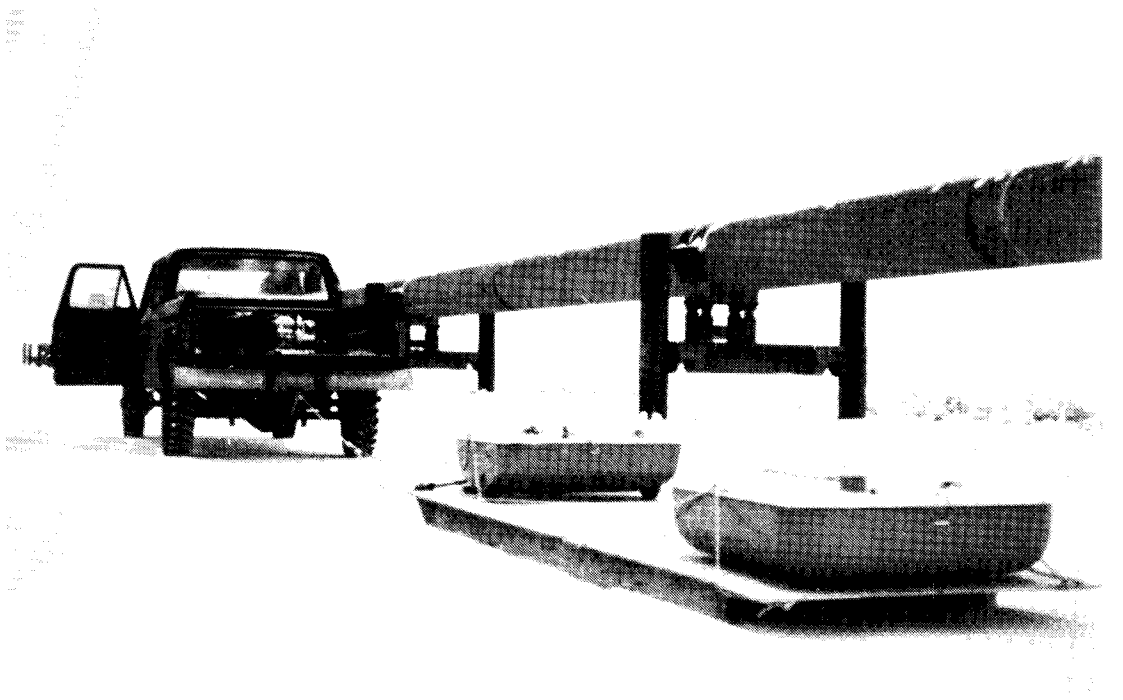

FIG. 1.6-11. Carrier-free dual radar used for probing massive ice in permafrost ground along the Alaska pipeline. (Courtesy A. Kovacs, U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire, and R. Morey, Morey Research Co., Nashua, New Hampshire.)

measure the propagation velocity in addition to the propagation time of the radar pulse to make full use of ground-probing radars.

The problem of measuring simultaneously propagation time and velocity was neatly solved by means of the *dual radar* (Kovacs and Morey, 1979b). Figure 1.6-11 shows such a dual radar used for profiling regions of massive ice in the permafrost ground along the Alaska pipeline. The usual transmitting/receiving antenna is augmented by a second antenna that receives only. A recording of the ground profile is shown in Fig. 1.6-12. The numbers 218 ... 240 on top of the illustration identify the vertical support structures of the pipeline; two of these support structures are visible in Fig. 1.6-11. Two holes had to be drilled for each structure. The top and bottom of the massive ice penetrated by the holes on the west side of the pipeline are marked by circles in Fig. 1.6-12, whereas the stars mark top and bottom of the ice for the drill holes on the east side. The radar profile and the drill-hole data correlate quite well, major discrepancies being due to the change of thickness of the ice over short distances; e.g., at the vertical support structure 224 the drill hole on the west side (circles) shows the ice about 6 m thick, whereas the drill hole on the east side—which was only 3 m away—showed no ice at all.

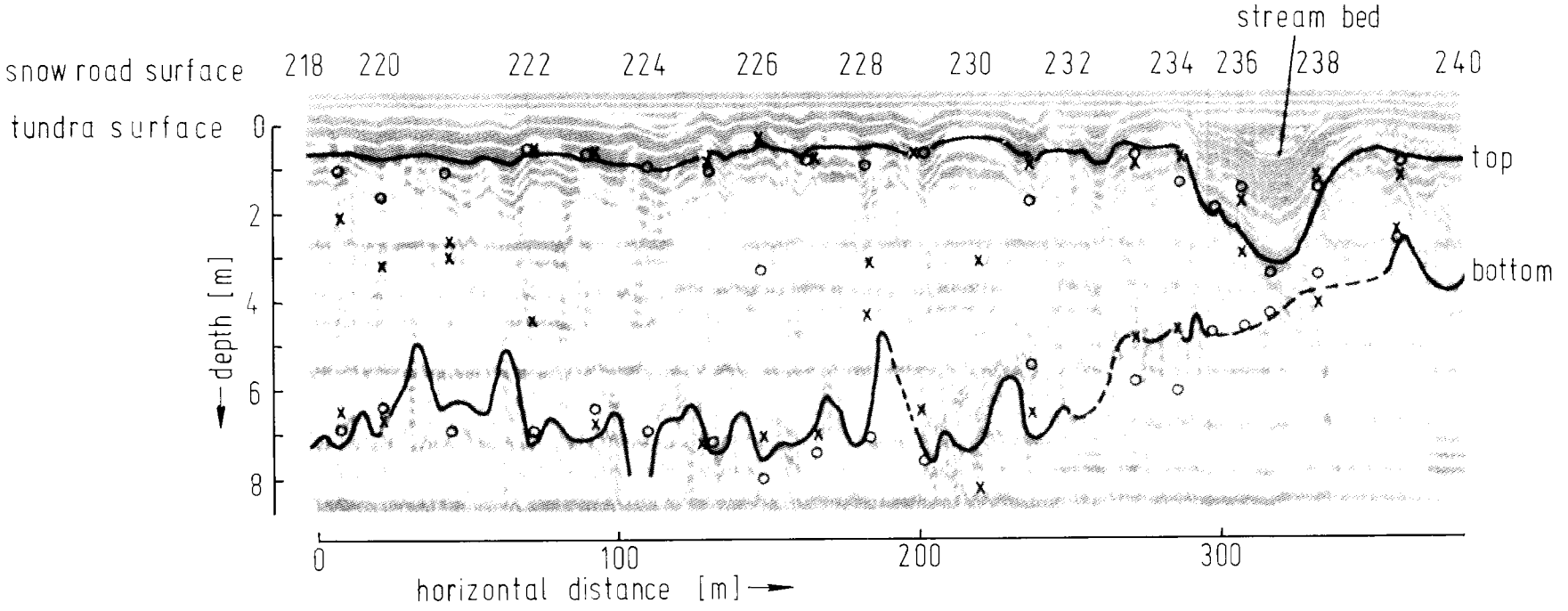

FIG. 1.6-12. Record of ground profile obtained by carrier-free dual radar and by drilling along the Alaska pipeline from vertical support structure 218 to 240. (Courtesy A. Kovacs, U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire, and R. Morey, Morey Research Co., Nashua, New Hampshire.)

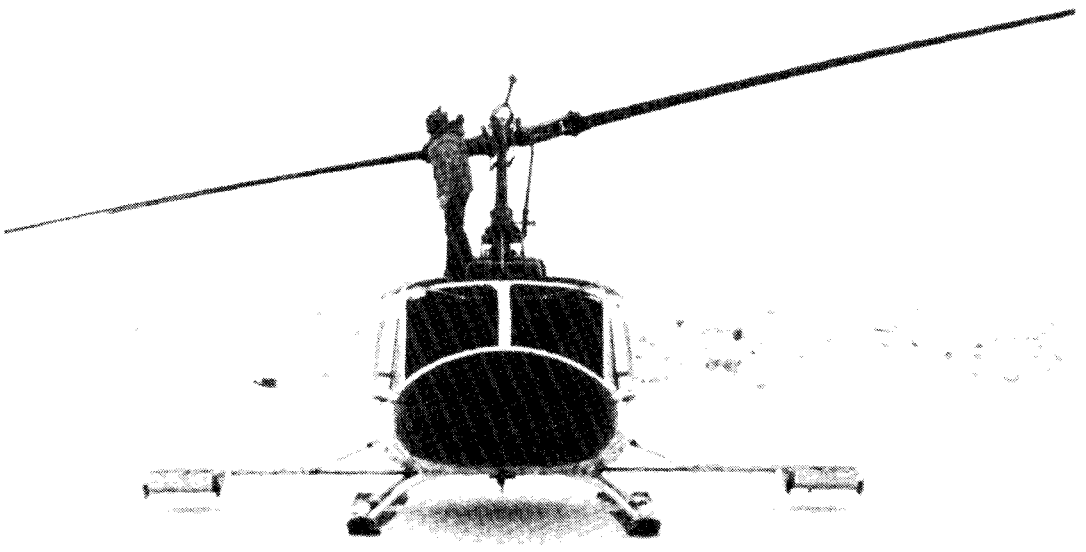

FIG. 1.6-13. Helicopter-mounted carrier-free dual radar used for surveys at Prudhoe Bay, Alaska. The man on the top of the helicopter is deicing the rotor. (Courtesy A. Kovacs, U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire.)

Figure 1.6-13 shows a carrier-free dual radar with two receiving antennas mounted on a helicopter. This radar was used for surveys at Prudhoe Bay, Alaska.

A carrier-free dual radar probing the Ross Ice Shelf near McMurdo Sound in Antarctica is shown in Fig. 1.6-14. The great distance between the two antennas indicates that probing of substantially greater depths than in Figs. 1.6-11 and 1.6-13 is attempted.

A carrier-free radar designed for ground-probing from tunnels of coal mines is shown in Fig. 1.6-15. This radar produces rather long pulses of 25 ns duration with a peak power of 2 MW. The long pulses and the high power permit a probing distance of about 60 m in sandstone and about 9 m in shale, which is frequently encountered in coal mines. This radar was built by J. C. Cook, whose paper of 1960 we mentioned as the beginning of the carrier-free into-the-ground radar (J. C. Cook, 1975).

The use of carrier-free radar as airport ground control sensors and simi-

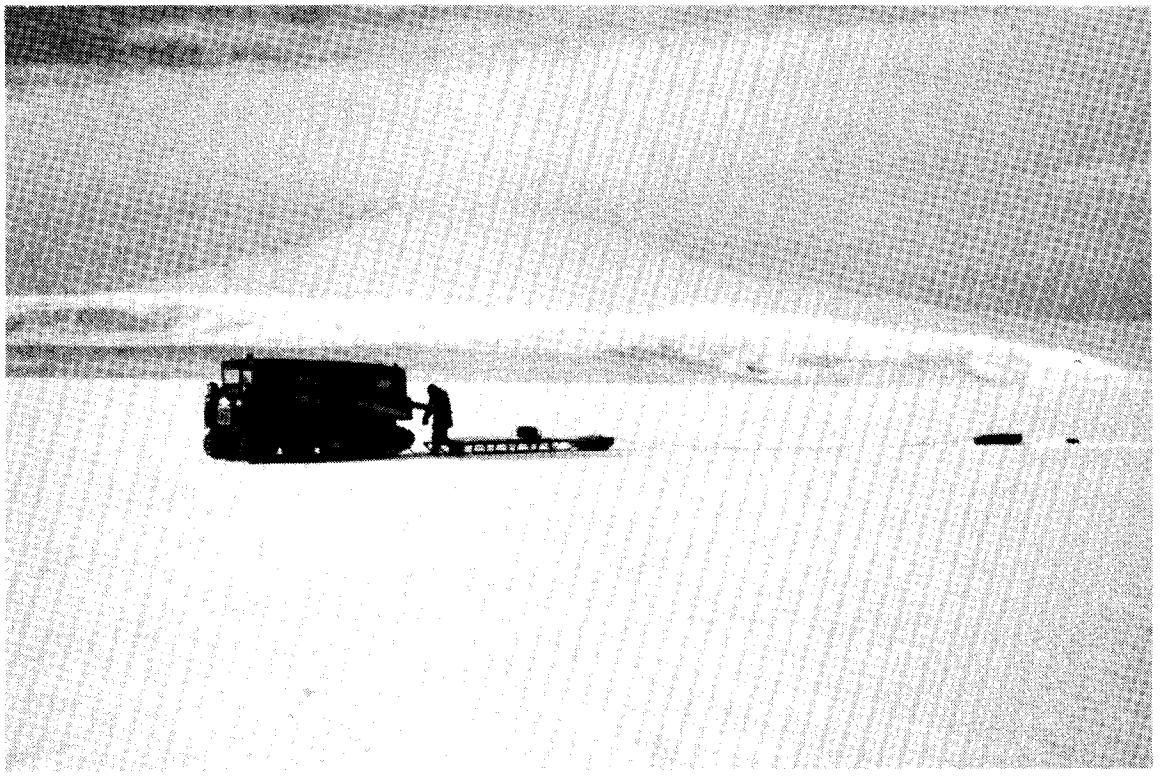

FIG. 1.6-14. Profiling of the Ross Ice Shelf at McMurdo Sound, Antarctica, with a carrier-free dual radar. (Courtesy A. Kovacs, U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire.)

lar applications were described by Bennett, Nicholson, and Ross (Ross, 1974; Nicholson and Ross, 1975; Bennett and Ross, 1978).

In Chapters 2–4, we will briefly discuss basic concepts of implementation of radio transmission with large relative bandwidth. However, most of this book is devoted to principles and applications; they come first. Technology will be developed once sufficiently good principles and applications become widely known.

Experience has shown that many engineers have difficulties with the technology of radio signals with large relative bandwidth, since most of our technological advancement is done in small steps, and any radical change is thus not part of their experience. This is typical for a mature industry. There were few small but many radical changes when radio transmission got started around 1900. To obtain some idea of what one should expect, let us look at technology from a broader point of view.

The relative bandwidth η defined by Eq. (1.4-1) can have any value in the range $1 \geqq \eta \geqq 0$. Our current technology is based on a theory for the limit $\eta \to 0$. Both theory and technology that apply to the whole range

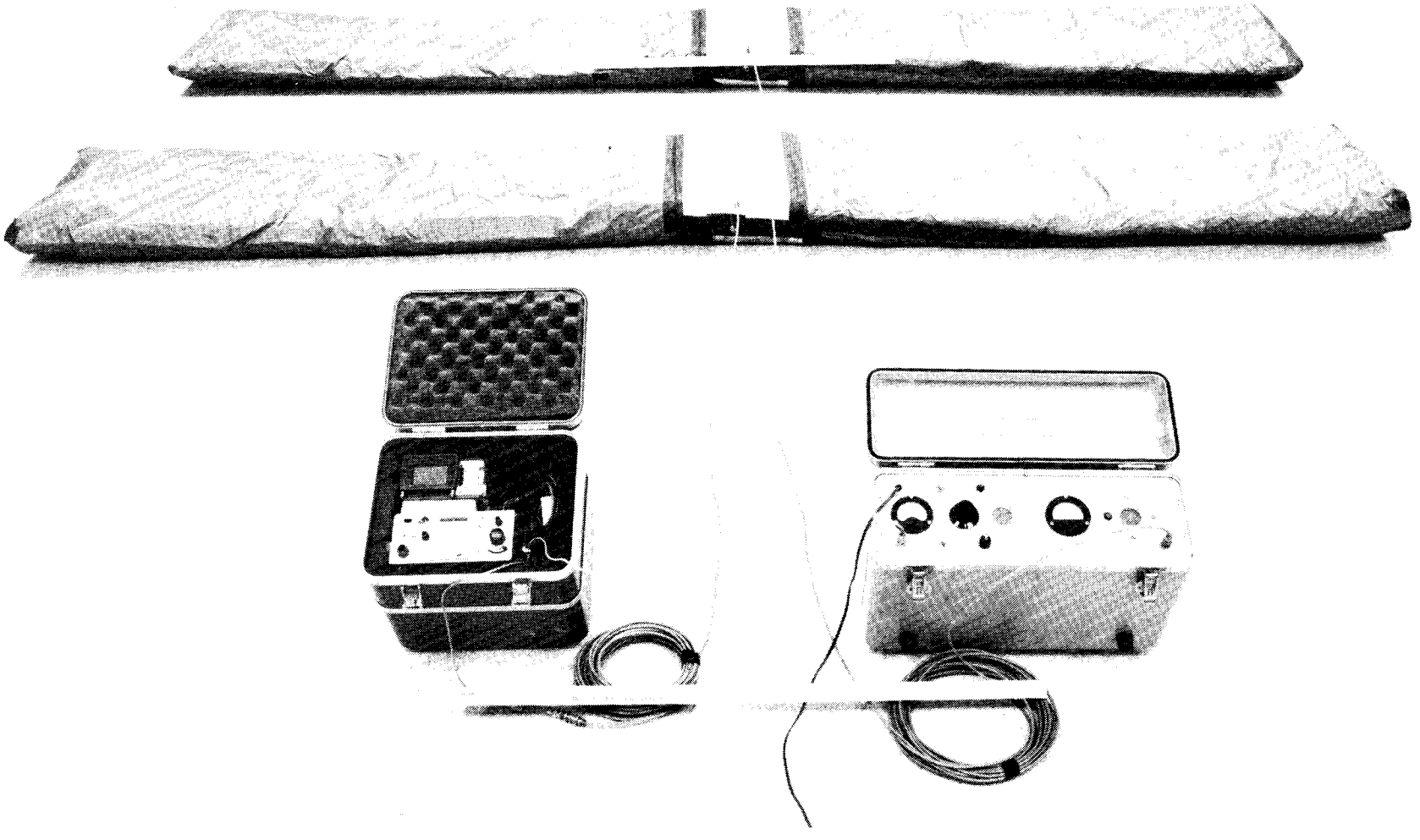

FIG. 1.6-15. Carrier-free radar using pulses of 25 ns duration and 2 MW peak power for probing in coal mines. The radiator on top is driven by a 10-kV thyratron (bottom right), the receptor feeds the receiver (bottom left) consisting of an amplifier, sampler, and oscilloscope. (Courtesy J. C. Cook, Teledyne-Geotech, Dallas, Texas.)

$1 \geqq \eta \geqq 0$ will have to be more general and more sophisticated than a theory and technology that apply to the limit $\eta \to 0$ only. Taking *Radar Handbook* (Skolnik, 1970) as a representative summary for the limit $\eta \to 0$ gives some idea of what one should reasonably expect for the general case $1 \geqq \eta \geqq 0$. It took 40 years of radar development, including several technology-advancing wars, before *Radar Handbook* could be compiled, and this indicates that a great deal of patience will be required before a comparable summary can be assembled for the general case.[1]

[1] As an example of an area of application that cannot be discussed in this book despite its potential, consider the problem of making an aircraft invisible to radar. At high frequencies, such as 35 or 94 GHz, one may use absorbing materials since the wavelength is in the range from 10 to 3 mm. At a frequency of 500 MHz, the wavelength is 60 cm, and the reduction of the backscattered wave by means of absorbing materials is hard to implement. This effect favors the use of low frequencies. On the other hand, the suppression of a backscattered wave by the radiation of an amplitude-reversed, but otherwise equal, wave is easy for a pure sinusoidal wave, but becomes increasingly difficult for signals with larger bandwidth. Furthermore, the larger the bandwidth of a signal the more information it can carry, which translates into a large radar signature. Both effects favor the use of large bandwidths. The radar most suited to overcome the problem of the invisible aircraft must thus use a large bandwidth and operate at low frequencies, which means its relative bandwidth should be close to 1.

Nonsinusoidal Waves for Radar and Radio Communication

HENNING F. HARMUTH

DEPARTMENT OF ELECTRICAL ENGINEERING
THE CATHOLIC UNIVERSITY OF AMERICA
WASHINGTON, D.C.

1981

ACADEMIC PRESS

A Subsidiary of Harcourt Brace Jovanovich, Publishers

New York London Toronto Sydney San Francisco