Antennas and Waveguides for Nonsinusoidal Waves

HENNING F. HARMUTH

Antennas and Waveguides for Nonsinusoidal Waves

HENNING F. HARMUTH

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

1984

ACADEMIC PRESS, INC.

(Harcourt Brace Jovanovich, Publishers)

Orlando San Diego New York London
Toronto Montreal Sydney Tokyo

ACADEMIC PRESS, INC.
Orlando, Florida 32887

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 63-12814

ISBN 0-12-014577-4

PRINTED IN THE UNITED STATES OF AMERICA

84 85 86 87 9 8 7 6 5 4 3 2 1

Antennas and Waveguides for Nonsinusoidal Waves

D563/07

Advances in Electronics and Electron Physics

Edited by

PETER W. HAWKES

Laboratoire d'Optique Electronique
du Centre National de la Recherche Scientifique
Toulouse, France

SUPPLEMENTS

1. Electroluminescence and Related Effects, 1963 HENRY F. IVEY
2. Optical Masers, 1964 GEORGE BIRNBAUM
3. Narrow Angle Electron Guns and Cathode Ray Tubes, 1968 HILARY MOSS
4. Electron Beam and Laser Beam Technology, 1968 L. MARTON AND A. B. EL-KAREH
5. Linear Ferrite Devices for Microwave Applications, 1968 WILHELM H. VON AULOCK AND CLIFFORD E. FAY
6. Electron Probe Microanalysis, 1969 A. J. TOUSIMIS AND L. MARTON, EDS.
7. Quadrupoles in Electron Lens Design, 1969 P. W. HAWKES
8. Charge Transfer Devices, 1975 CARLO H. SÉQUIN AND MICHAEL F. TOMPSETT
9. Sequency Theory: Foundations and Applications, 1977 HENNING F. HARMUTH
10. Computer Techniques for Image Processing in Electron Microscopy, 1978 W. O. SAXTON
11. Acoustic Imaging with Electronic Circuits, 1979 HENNING F. HARMUTH
12. Image Transmission Techniques, 1979 WILLIAM K. PRATT, ED.
13. Applied Charged Particle Optics, 1980, 1983 (in three parts) A. SEPTIER, ED.
14. Nonsinusoidal Waves for Radar and Radio Communication, 1981 HENNING F. HARMUTH

Foreword

Professor Henning Harmuth is no stranger to readers of these *Advances*. This further contribution to the field of nonsinusoidal waves, which he has made peculiarly his own, is very welcome for it takes the subject from the realm of theory into the much more difficult domain of the practical design of antennas and waveguides for these waves. The author has gone to considerable trouble to make this unfamiliar and relatively difficult material accessible to the scientists and engineers for whom these new techniques are destined, and I have no doubt that this text will become essential reading as the methods gain a wider audience.

P. W. HAWKES

Preface

Over the last few years much has been published about the principles and applications of electromagnetic waves with large relative bandwidth, or nonsinusoidal waves for short. The next step is the development of the technology for the implementation of these applications.

It is generally agreed that the antennas pose the most difficult technological problem, since circuits for nanosecond and subnanosecond pulses have existed for some time, and the possible time variation of electromagnetic waves has been known since Maxwell. A very similar problem was faced in the development of the technology for the transmission of sinusoidal waves around 1900. It was overcome by the resonating dipole antenna. Marconi is usually credited today with the introduction of this antenna on the strength of his patents, but the issue created much controversy in its time.

The existing antenna theory was developed since 1900 with sinusoidal waves in mind, even though the connection of antennas with sinusoidal waves and their parameter frequency is sometimes stated in a negation by calling them broadband or frequency-independent antennas. What is clearly needed for nonsinusoidal waves is an antenna type that is about as fundamental as the resonating dipole. The large-current radiator and the closed-loop sensor appear to be such a fundamental type.

The mathematical requirements for a theory of antennas for nonsinusoidal waves are substantially higher than for the theory of antennas for sinusoidal waves. The reason is that sinusoidal waves are particular solutions of Maxwell's equations, while nonsinusoidal waves are more general solutions. The separation of variables by Bernoulli's product method is replaced by vector analysis as the basic mathematical tool. This is done in this book in considerable detail in order to make the methods useful for the electrical engineer.

The very fact that nonsinusoidal waves represent more general solutions of Maxwell's equations than sinusoidal waves makes one expect results that go qualitatively beyond what sinusoidal waves can do. The investigation of radiation patterns of antenna arrays fully bears out this expectation.

A theory of waveguides for sinusoidal waves has existed for about ninety years. Its generalization for nonsinusoidal waves logically calls again for vector analysis, but the mathematically simpler—even though less powerful—way via Fourier series and transform is chosen here. This simplified approach appears justified since we are already familiar with the distortion-free transmission of nonsinusoidal waves by coaxial waveguides or cables from the theory of the distortion-free transmission line.

Some of the results presented here were originally published in the *IEEE Transactions on Electromagnetic Compatibility*. The author wants to thank the editor Richard B. Schulz for his support in overcoming the usual opposition to the publication of new ideas. Furthermore, he wants to thank the IEEE Electromagnetic Compatibility Society in general and its members A. T. Adams, H. K. Mertel, G. R. Redinbo, G. F. Sandy, H. M. Schlicke, and L. W. Thomas in particular for more than a decade of support. Chang Tong of Qinghua University in Beijing; Fan Changxin and Hu Zheng of Northwest Telecommunication Engineering Institute in Xi'an, Shensi; Xie Chufang of Chengdu Institute of Radio Engineering, Sichuan; Zhang Qishan of Beijing Institute of Aeronautics and Astronautics; and B. A. Austin, H. E. Hanrahan, C. F. Landy, J. P. Reinders, M. G. Rodd—all of University of the Witwatersrand, Johannesburg, South Africa—and T. W. Cole of University of Sydney, Australia, have helped the author in various ways. This help is gratefully acknowledged.

The experimental work on antennas for radar and radio transmission reported in Section 3.11 was supported by Geophysical Survey Systems Inc. of Hudson, New Hampshire. But science blooms with a hundred blossoms, and the author's work was further supported by the Government of the Peoples Republic of China, the University of the Witwatersrand, South Africa, and the University of Sydney, Australia. There is no greater gift for the research engineer than the opportunity to maintain flourishing experimental work.

The author wants to take the opportunity to direct attention to the books on radio transmission by D. D. Klovsky and co-workers of Kuibyshev, Soviet Union (1969, 1975, 1976, 1984). The lack of English editions is delaying dissemination of their results just as it happened with Kotel'nikov's work a generation ago.

Contents*

*Equations are numbered consecutively within each of Sections 1.1–6.6. Reference to an equation in a different section is made by writing the number of the section in front of the number of the equation, e.g., Eq. (6.1-54) for Eq. (54) in Section 6.1.

Illustrations and tables are numbered consecutively within each section, with the number of the section given first, e.g., Fig. 1.6-2, Table. 3.3-1.

References are characterized by the name of the author(s), the year of publication, and a lowercase Latin letter if more than one reference by the same author(s) is listed for that year.

1 Introduction

1.1 Basic Concepts

Theory and technology for radio transmission have been solidly based on the concept of almost sinusoidal waves since the beginning of radio engineering around 1900. The reason that the theory was directed toward sinusoidal waves was due to the difficulty of finding solutions of Maxwell's equations without using Bernoulli's product method. Although this method permits the solution of partial differential equations via ordinary differential equations, only particular solutions, rather than the general solution, are obtained, and these are in terms of sinusoidal functions. The general solution of Maxwell's equations for the radiation of the Hertzian electric dipole became available in English more than half a century ago when the first translation of the renowned book *Theorie der Elektrizität* was published.[1] This solution yields the electric and magnetic field strengths as functions of the dipole moment, or the dipole current, without the usual assumption of a sinusoidal time variation of this moment or current.

The engineering literature has almost completely ignored this general solution. Textbooks, scientific books on antennas, and nearly all journal publications assume a sinusoidal time variation for antenna currents and the waves produced. The development of technology was, of course, handicapped by this restriction of the widely read literature. Furthermore, the discovery

[1] This book was first published by August Föppl in 1894 under the title *Einführung in die Maxwellsche Theorie*. Max Abraham revised it in 1904, changed the title to its current form *Theorie der Elektrizität*, and published the second to seventh editions of the first volume. In 1930, Richard Becker took over and published the eighth to fifteenth editions. The sixteenth edition was in preparation when Becker died in 1955 and Fritz Sauter continued the work. The first as well as the second English edition (Abraham and Becker, 1932, Part III, Chapter X, Sect. 11; 1950, Part III, Chapter X, Sect. 10) contain the general solution of the Hertzian electric dipole in component form H_α, E_r, E_ϑ. The third English edition (Becker, 1964, Chapter DIII, Sect. 67) contains it in vector form. The field strengths **E** and **H** produced by a point charge having a velocity **v** with arbitrary time variation may already be found in vector form in the second German edition (Abraham, 1905, Vol. 2, Sect. 13); the field strengths of the Hertzian electric dipole may readily be derived from these equations. Abraham credits K. Schwarzschild in a footnote on page 97 for having published the derivation in 1903 (Schwarzschild, 1903). The solution of Maxwell's equations for the Hertzian electric dipole with a sinusoidal current is due to Hertz (1889, 1893).

that sinusoidal currents and voltages could resonate with *LC* circuits and certain antenna structures made selective radio transmission practically possible, but it also created the widely held belief that resonance is a phenomenon peculiar to sinusoidal functions. More generally, the theory based on sinusoidal functions led to a technology for sinusoidal functions which in turn stimulated the development of more theory for sinusoidal functions. Looking back, it is difficult to say with certainty what broke this circle. It was probably the development of semiconductors and the renewed interest in digital circuits, now based on electronic rather than electromechanic components, that impressed on our thinking that sinusoidal functions could not always be used.

Since information is always transmitted by nonsinusoidal[2] functions, we have to be more specific in what we mean by nonsinusoidal functions. We use the relative bandwidth η as a more technical term. The relative bandwidth is defined by

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

where f_H is the highest and f_L the lowest frequency of interest. This definition can be applied to the amplitude modulation of a sinusoidal carrier with frequency f_c by a signal with bandwidth Δf. One obtains $f_H = f_c + \Delta f$, $f_L = f_c - \Delta f$, and

$$\eta = \Delta f / f_c \tag{2}$$

which is the usual form of the relative bandwidth found in textbooks. Let us observe that Eq. (1) is more general than Eq. (2) since no reference is made to a carrier frequency f_c in Eq. (1).

For a sinusoidal function, we have $f_H = f_L$ and $\eta = 0$. A typical radio or radar signal yields $\eta \doteqdot 0.01$; such a signal is no longer a pure sinusoidal function, but it still looks very sinusoidal. As η increases, the similarity to a sinusoid decreases, and no semblance to a sinusoidal function remains as the relative bandwidth approaches its upper limit $\eta = 1$. When we thus talk about "sinusoidal" or, better yet, almost sinusoidal signals, we mean the limit $\eta \to 0$, whereas a nonsinusoidal signal means a relative bandwidth anywhere in the interval $0 < \eta \leq 1$ and "significantly" larger than zero.

---

[2] More precisely, information is transmitted at a rate larger than 0 only by time functions $f(t)$ that are not analytic in the whole interval $-\infty < t < +\infty$. If an analytic function is known for a particular time t_0, it is known for all times, and no information is gained by transmitting it for other times. Functions that are analytic only in a finite interval do transmit information at a rate larger than 0. Hence, short and long sinusoidal *pulses* produced by a telegrapher's key transmit information at a rate larger than 0 but the infinitely long periodic sinusoidal function does not. Besides analytic functions, periodic functions that may or may not be analytic transmit information at the rate 0.

One must be careful to observe that the relative bandwidth is not a precise concept. Any (time) signal in the real world has a beginning and an end; it has a finite energy, a finite amplitude, and a finite change ΔA of its amplitude requires a finite time ΔT. The Fourier transform differs almost everywhere from zero in the interval $0 \leqq f < \infty$. Hence, f_L and f_H cannot be the frequencies below and above which the Fourier transform is zero everywhere. We must arbitrarily define what we mean by highest and lowest frequencies of interest; e.g., that no more than 1% of the signal energy is below f_L and above f_H. This is an unavoidable difficulty arising from the use of concepts derived from sinusoidal functions and Fourier analysis. It is caused by the implied representation of a signal of finite duration—a nonanalytic function—by a superposition of sinusoidal functions of infinite duration, which are analytic in the whole interval $-\infty < t < +\infty$. One could avoid the problem by doing away with the Fourier transform[3] and concepts such as frequency and frequency bandwidth based on it, but few radio engineers would appreciate mathematical precision at such a price. Hence, we shall use the concepts of Fourier analysis whenever we can and avoid their use only in specifically indicated cases.

1.2 Transmitter and Receiver for Nonsinusoidal Waves

Since transmitters and receivers for nonsinusoidal signals differ considerably from conventional ones, we first discuss block diagrams of such equipment before we turn to the more detailed investigation of antennas and waveguides for such equipment. Figure 1.2-1 shows the block diagram of a carrier-free radar. The transmitter consists of a signal generator SGE that

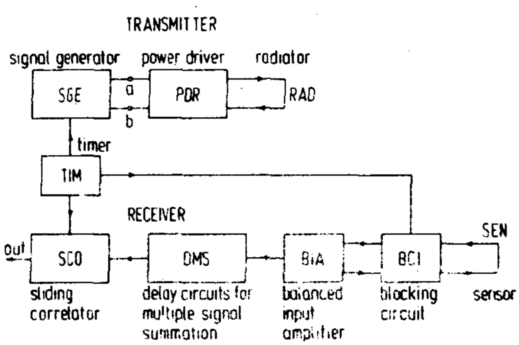

FIG. 1.2-1. Block diagram of a carrier-free radar.

[3] The Fourier series is not affected since it requires a finite interval only. The usual periodic continuation outside this finite interval is not a necessity.

feeds a current via a power driver PDR to the radiator RAD. Typical time variations of the voltages at the points a and b, as well as of the current in the radiator, are shown in lines a and b of Fig. 1.2-2. These sawtooth currents in the radiator produce electric and magnetic field strengths with the time variation shown in line c. The short positive and the long negative pulses have the same area $E\,\Delta T$. The energy of the positive pulse is thus proportionate to $E^2\,\Delta T$, whereas that of the negative pulse is proportionate to $(E\,\Delta T)^2/(T - \Delta T)$. For $\Delta T \ll T$, practically all the energy is thus in the short positive pulses. For this reason, we can often ignore the long negative pulses, even though they assure that the radiated signal has no dc component.

Let the positive pulses of Fig. 1.2-2, line c, be returned by a pointlike scatterer or a radar reflector. The choice of these two targets assures that the returned pulse has the same time variation as the radiated pulse. Other targets will cause a change that is sometimes called a distortion, more properly, a radar signature, since the change gives us information about the shape and materials of the target.

The returned pulse is received in Fig. 1.2-1 by a sensor and passes to a blocking circuit. If the radiator and the sensor use the same antenna structure, the blocking circuit becomes a duplexer. Commercially available isolators and circulators, which can implement duplexers, have relative bandwidths up to about $\eta = 0.67$. This is a very large value for use in conventional radar, but insufficient for carrier-free radar. A practical duplexer for carrier-free radar was described by Morey (1974) and is used in many of the existing carrier-free radars. Other designs, particularly those using separate antennas for radiation and reception, implement the blocking circuits by switches that disconnect the sensor from the balanced input amplifier BIA and short the input terminals of the amplifier during radiation. The balanced input amplifier suppresses the common mode of the sensor and amplifies the differential mode.

So far, transmitter and receiver have not differed much from conventional equipment; indeed, if the signal generator SGE in Fig. 1.2-1 is replaced by a

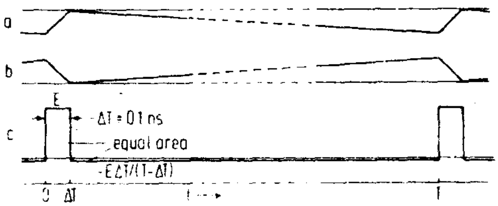

FIG. 1.2-2. Time diagram for the carrier-free radar of Fig. 1.2-1 when a simple rectangular pulse is used on the radio link.

sinusoidal signal generator, one could not tell the difference. This similarity ends at the output terminal of the balanced input amplifier, since we now need selective circuits that enhance the transmitted signal over noise and all other signals.

The circuits providing selectivity in Fig. 1.2-1 are the two blocks denoted as delay circuits for multiple-signal summation DMS and sliding correlator SCO. Let equal signals be received at the times $-60T$, $-50T$, ... 0, as shown in line *a* of Fig. 1.2-3. The time variation of these signals is of no interest yet; only their time position is needed. Let this sequence of seven equal signals be applied to the feedback-delay circuit in Fig. 1.2-3. The signal is fed through the hybrid coupler HYC1 to the delay circuit DEL with a delay of $10T$. From the output of the delay circuit it is fed back via the hybrid coupler HYC2, the amplifier AMP, and the variable attenuator VAT to the hybrid coupler HYC1, where it is added to the next incoming signal.

Let the variable attenuator VAT be set so that the gain in the feedback loop is exactly 1. At time $-60T$ signal 1 of line *a* with amplitude A arrives at the terminal *a* of HYC1. After passing through the feedback-delay loop, it arrives at the time $-50T$ with amplitude A at the terminal *c* of HYC1. At the same time, signal 2 with amplitude A arrives at terminal *a*; the two are summed to a signal with equal time variation but with amplitude $2A$ at terminal *b*. This signal runs again through the feedback-delay loop, arrives at the time $-40T$ at terminal *c* of HYC1, and is added to signal 3 arriving at the same time to yield amplitude $3A$ at terminal *b*. The signal in the feedback loop increases as shown in line *b* of Fig. 1.2-3. After signal 7 has been received, the amplitude of the signal in the feedback loop is $7A$, and it remains constant from then on. Hybrid coupler HYC2 makes it possible to feed the enhanced signal out for further use.

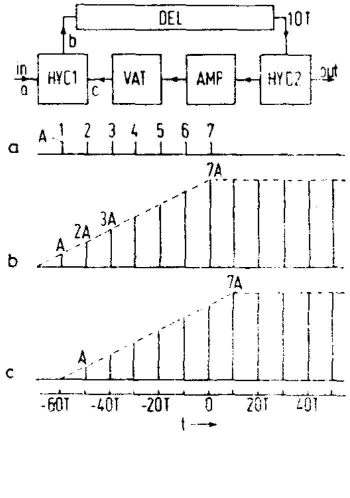

FIG. 1.2-3. Resonant circuit using the feedback delay principle and typical voltages at the points *a*–*c*: HYC, hybrid coupler; VAT, variable attenuator; AMP, amplifier; DEL, delay circuit.

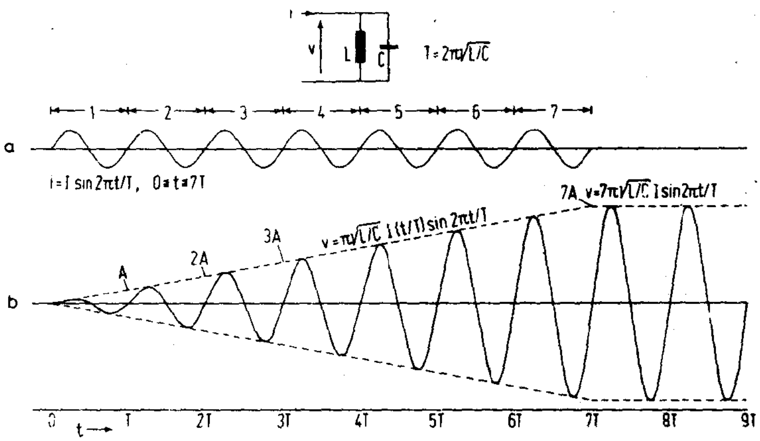

FIG. 1.2-4. Parallel LC circuit with a current flowing into it that has the time variation of a sinusoidal pulse with seven cycles (line *a*), and the resulting voltage (line *b*).

The linear increase of the recycling signal in the time interval $-60T \leq t \leq 0$ and the constant amplitude for $t > 0$ reminds one of an *LC* circuit if a sinusoidal pulse is applied. Figure 1.2-4 shows such a resonant circuit. Let a current $i = I \sin 2\pi t/T$ with seven cycles as shown in line *a* be fed into the circuit. The voltage v produced by this current is shown in line *b*. For a lossless *LC* circuit, the voltage v will keep oscillating after the current i has stopped at the time $t = 7T$. Losses will make these oscillations decay, just as a gain smaller than 1 in the feedback circuit of Fig. 1.2-3 will cause a decay of the amplitudes in line *b* for $t > 0$.

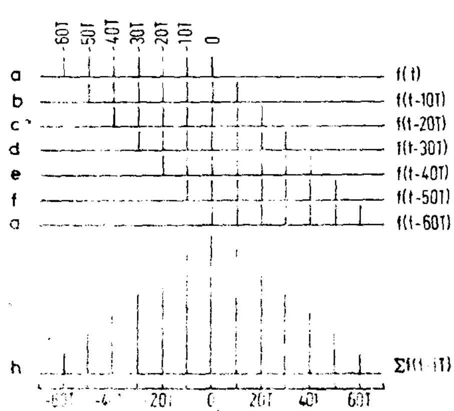

FIG. 1.2-5. Principle of a delay resonant circuit producing linearly increasing and decreasing output voltages.

A slow decay of the signals in line b of Figs. 1.2-3 and 1.2-4 after the received signals in line a have terminated is a drawback, since it limits the possible time resolution. It would be better if the decaying part of the output signal could be made equal to the linearly increasing part. To derive a respective circuit, we repeat the signal of line a of Fig. 1.2-3 in line a of Fig. 1.2-5. We denote this signal as $f(t)$ and produce the additional delayed signals $f(t - 10T)$, $f(t - 20T), \ldots, f(t - 60T)$ shown in lines b–g. All seven signals $f(t)$ to $f(t - 60T)$ are summed in line h of Fig. 1.2-5. The desired linear increase and delay are obtained. A circuit that implements this process is shown in Fig. 1.2-6a. A hybrid coupler HYC1 splits the received signal into seven signals that are delayed according to lines a–g in Fig. 1.2-5 and summed by hybrid coupler HYC2.

Consider again a sinusoidal pulse with seven cycles as shown in Fig. 1.2-7, line a. If this signal is applied to terminal a of hybrid coupler HYC1 in Fig. 1.2-3, and the delay in the delay line DEL is reduced from $10T$ to T, one obtains at terminal b of hybrid coupler HYC1 the voltage of line b in Fig. 1.2-7. We note that this voltage is very similar to that in line b of Fig. 1.2-4, but the voltages are not equal. Now let the signal of Fig. 1.2-7, line a, be fed into the circuit of Fig. 1.2-6a, again with all delays reduced to one-tenth. The output signal is shown in line c of Fig. 1.2-7. We now have the same linear increase and decrease as in Fig. 1.2-5, line h.[4]

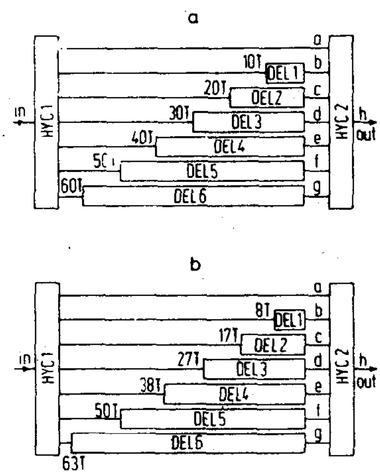

FIG. 1.2-6. Implementation of the delay resonant circuit principle for a signal according to Fig. 1.2-5 (a) and a more sophisticated signal (b).

[4] The delay circuit of Fig. 1.2-6a is superior in this respect to a resonant circuit like the one in Fig. 1.2-4. However, the delay circuit will enhance any sinusoidal pulse with a period T/n and nm cycles essentially equally, where $n = 1, 2, \ldots,$ and $m = 7$ in our case. The resonant circuit of Fig. 1.2-4, on the other hand, will only enhance sinusoidal pulses with period T.