Introduction

The Dogma of the Circle

The Greek philosopher Plato is credited with the introduction of the dogma of the circle. It was expressed by Claudius Ptolemy as follows:¹ “We believe that the object which the astronomer must strive to achieve is this: to demonstrate that all the phenomena in the sky are produced by uniform and circular motions.” The astronomical observations of the Greeks were accurate enough to show that the planets were not moving on circles or on surfaces of spheres, regardless of whether one assumed the Earth or the Sun as the center of motion. Eudoxus, a disciple of Plato, used a superposition of rotating spheres in an attempt to reconcile the observed data with the dogma of the circle. Four spheres were needed for each one of the five known planets,² three each for the Sun and the Moon, and one for the fixed stars. These 27 spheres proved unsatisfactory. Aristotle reduced the discrepancies between theory and observation by using 54 spheres.

Claudius Ptolemy replaced the spheres by circles. The five planets, the Sun, and the Moon moved around the Earth on primary circles called deferents. Superimposed on each deferent was a secondary circle, called epicycle, as shown in Fig. 0-1. Another epicycle was superimposed on this first epicycle, and so on. In modern language, we would say that the orbits were represented by a superposition of circles. Ptolemy used 36 circles to

![Diagram of epicycle and deferent](image)

Fig. 0-1. Superposition of circles in astronomy and in electrical communications.

¹ A detailed account of the dogma of the circle in astronomy is given by Koestler (1959). German translations of the most pertinent Greek and Latin texts are collected in a book by Zinner (1951).
² Mercury, Venus, Mars, Jupiter, and Saturn are visible to the unaided eye.
represent the orbits of the Sun, the Moon, and the five planets. This was not quite enough to fit the observed data, and better representations using more circles were subsequently worked out. Nicholas Copernicus moved the center of motion from the vicinity of the Earth to the vicinity of the Sun, but he retained the representation of orbits by a superposition of circles. The orbit of Mercury required eleven circles, Venus and Earth nine circles each, the Moon four circles, and the remaining planets five circles each. This adds up to 48 circles.

Johannes Kepler put an end to the superposition of circles in 1609, when he showed in his book “Astronomia Nova” that elliptical orbits fitted the observed data better and simpler.

It is generally believed that Kepler ended the dogma of the circle, but this is not so. The circle disappeared from astronomy, but it reappeared in other fields of science in disguise. In electrical engineering and a good part of physics, we meet the old circle under the new name of exponential function $e^{j\omega t}$ or unit circle in the complex plane. Anyone with the usual background of electrical communications will interpret Fig. 0-1 not as a superposition of a deferent and an epicycle but as a superposition of two sinusoidal oscillations $I_1 \exp(i\omega_1 t)$ and $I_2 \exp(i\omega_2 t)$ using complex notation.\(^1\) Indeed, Fig. 0-1 is a standard illustration for single sideband modulation of a sinusoidal carrier by a sinusoidal signal. Speaking more generally, the superposition of circles by Ptolemy and Copernicus became the Fourier series expansion in complex notation.

The expression character group of the topologic group of real numbers does not seem to have anything to do with the circle, but its mathematical notation \(\{e^{j2\pi x}\}\) reveals the truth. This character group implies the topology of the continuum for space–time, which in turn permits the use of differential calculus for functions of space and time. Considering the universal use of differential calculus in physics, one cannot help but suspect that the circle influences physics today as much as it once influenced astronomy.

Finding and studying hidden remnants of the dogma of the circle is the purpose of sequency theory.

Let us observe that the deferents and epicycles of Ptolemy represented orbits well enough to get Vasco da Gama to India, Christopher Columbus to America, and one of Ferdinand Magellan’s ships around the world. It was a sufficiently good theory for many practical purposes, but its finer details always indicated that something was not quite right. Similarly, the exponential function or the sine–cosine functions in communications have proved to be

\(^{1}\) $I_1 \exp(i\omega_1 t)$ and $I_2 \exp(i\omega_2 t)$ are called vectors in the older literature (Cherry, 1949; Cuccia, 1952) and phasors in the newer literature (Van Valkenburg, 1964; Taub and Schilling, 1971).
THE CIRCULAR FUNCTIONS IN COMMUNICATIONS

perfectly good for many applications, but it is generally known that something is not quite right. All real signals have an infinite frequency bandwidth, filters yield an output voltage before an input voltage is applied, etc. We use experience and common sense to correct for such deficiencies of the theory; but a correct theory would not need corrections, and the known need for corrections may be like the tip of the iceberg. Turning to the character group of the topologic group of real numbers, there can be no doubt about the success of differential calculus in physics; but it is unsatisfactory to talk about what is happening at a point $x$ and at another point $x + dx$, if we cannot make measurements at two points having a distance $dx$ from each other.

**THE CIRCLE AND THE CIRCULAR FUNCTIONS IN COMMUNICATIONS**

The unit circle in the complex plane, $e^{i\omega t} = \cos \omega t + i \sin \omega t$, and its decomposition into circular functions play a dominant role in electrical communications and physics. Whenever one uses the term frequency, one refers implicitly to these functions. Let us see how this dominant role came about and where its limitations are.

During the 19th century, the most important functions for communications were the block pulses shown in Fig. 0-2. Voltage and current pulses could be generated by mechanical switches, amplified by relays, and detected by a variety of magnetomechanical devices. Sine–cosine functions and the exponential function were well known and so was Fourier analysis, although

![Diagram of sine and cosine functions, Walsh functions, and block pulses.](image-url)

**Fig. 0-2.** Sine–cosine functions, Walsh functions, and block pulses.
in a somewhat rudimentary form. Almost no practical use could be made of this knowledge with the technology available at that time. Heinrich Hertz used the exponential function to obtain his famous solution of Maxwell’s equations for dipole radiation, but he was never able to produce sinusoidal waves. His experiments were done with what we would call colored noise today (Hertz, 1889). Alexander Graham Bell tried to develop telegraphy multiplex equipment using sinusoidal functions, but failed because he could not produce voltages with sinusoidal time variation. His voltages were square waves, while his receivers resonated with sine waves. Two results of this work were the introduction of the word *sinusoidal* into communications engineering and the discovery of voice transmission by electricity. Bell’s telegraphy transmitter decomposed voice into square waves and the receiver recomposed it from the square waves. Hence, the decomposition of voice into square waves predates the decomposition into sinusoidal waves by many decades (Bell, 1876; Marland, 1964).

Telegraphy equipment using orthogonal sine–cosine functions according to Bell’s concept was successfully developed more than seventy years later under names like Kineplex, Rectiplex, and Digiplex.

The first practical use of sinusoidal functions came toward the end of the 19th century with the development of capacitors in a useful form. Capacitors in the form of metallic spheres and Leyden jars had existed for a long time, but their capacitance was small and their physical structure inconvenient. The implementation of inductances through the use of coils had been known long before. Practical resonant circuits for the separation of sinusoidal electromagnetic waves with different frequencies could thus be built around the turn of the century. Low-pass and band-pass filters using coils and capacitors were introduced in 1915, and a large new field for the application of sinusoidal functions was opened. Speaking more generally, the usefulness of sinusoidal functions in communications is intimately related to the availability of linear, time-invariant circuit components in a practical form.

On the theoretical level, the use of sinusoidal functions is strongly favored by differential calculus, and our concept of the topology of space–time derived from it. This theoretical basis is discussed in some detail in the sections of this book devoted to physics.

The first indication that a theory of communications based on sine–cosine functions would eventually prove unsatisfactory comes from the importance of linear, time-invariant circuit components and circuits for these functions. One cannot transmit information if everything is (time) invariant. The telegrapher’s key, the microphone, and the amplitude modulator are linear but time-\textit{variable} devices. Making them time \textit{invariant} by not operating the key, not speaking into the microphone, or not feeding a time-variable modulating voltage into the modulator puts an end to the transmission of infor-
mation. The requirement for time variability for information transmission holds quite generally. An atom with all electrons in certain quantum states transmits no information. A photon is emitted by a change of quantum state, and this time variation provides information about the energy difference of the initial and final state.

Sine-cosine functions are obtained as the eigenfunctions of systems described by linear differential equations with constant coefficients. Hence, sine-cosine functions are most convenient as long as one may ignore the time variability. Increasing sophistication forces one to use equations (not necessarily differential equations) with variable coefficients; their eigenfunctions are no longer sinusoidal functions.

We have so far discussed time variability. Space variability is a straightforward extension. Sinusoidal functions have slanted our thinking heavily toward time signals. As an example, consider a tunable generator for sinusoidal functions. All the commercially available ones are generators for time-variable sinusoidal functions. Indeed, it is not only impossible to buy generators for space-variable sinusoidal functions but it is rather difficult even to imagine a generator that can be tuned, e.g., from 20 to 20,000 cycles per meter. Most textbooks do not mention space signals. Publications on filters are almost exclusively concerned with time signals.

A simple example of a space signal with two variables is a black-and-white photograph that has various shades of gray as a function of \( x \) and \( y \) in cartesian coordinates or of \( r \) and \( \phi \) in polar coordinates. A television signal is a function of two space variables and the time variable. The ubiquity of TV signals makes it safe to conclude that most transmitted information does not consist of functions of the time variable only. Beyond TV, most of the information received by us comes through the eyes and not the ears.

Why then do we hear so little about space signals and filters for space signals? One reason is that the concept of time invariance, meaning that something has always been as it is now and will always remain so, is acceptable to our thinking although we know it is unrealistic. Space invariance, on the other hand, is so unrealistic that we cannot accept it. A television image clearly has a left and a right edge, a top and a bottom, while the finite extension in time is much less obvious. There are \( 30 \times 3600 = 108,000 \) images as a sequence of time per hour according to the U.S. standard, but none of the more widely used TV systems has more than some five or six hundred space points in the \( x \) and \( y \) direction. Hence, a theory of space filters must begin with space-variable filters and cannot consider space-invariant filters as a starting point. A second reason for not hearing much about filters for space signals is that filters for time signals are overwhelmingly implemented by inductances and capacitances, but this technology is not applicable to filters for space signals.
We may answer at this point a question that has often been raised: Why should one use nonsinusoidal functions when sinusoidal functions have proved to be so good for theoretical investigations and practical applications? The answer is that there are certain uses for which sinusoidal functions are good. These are the uses that have been developed during the last eighty years. There are other uses for which sinusoidal functions are not good and which therefore have not been developed and are not found in our textbooks. Spatial electric filters were not derived from sinusoidal functions. Television scanners based on sine–cosine functions were never developed. Sinusoidal electromagnetic waves cannot be used to discriminate between a reflector and a scatterer, or between a conducting and a nonconducting scatterer. Both effects are of great interest in radar, but their very existence escaped our attention as long as we were restricted to thinking in terms of sinusoidal waves. Multipath transmission is known to lead to signal cancellation due to interference fading, but only for sinusoidal and other polarity-symmetric waves. Several more effects of electromagnetic waves have been found that are so obscure for sinusoidal waves that they were never noticed.

Let us turn to the third basic reason for going beyond sine–cosine functions: the convergence of Fourier series and Fourier transform. Any practical signal can be approximated by the Fourier series or the transform in the sense of a vanishing mean-square error. Mean-square convergence implies that the energy of a signal is the same as that of a superposition of sine–cosine functions approximating it. This preservation of energy is certainly necessary, but it is not sufficient for the transmission of information. For an explanation of this statement refer to Fig. 0-2. The Walsh functions shown there are characterized by the location of the zero crossings or sign changes. The constant sections between the zero crossings can always be filled in; they convey no information. A Fourier series or transform of these Walsh functions converges everywhere except at the zero crossings. The divergence at these “jumps” is so well known that it received its own name, Gibbs phenomenon. Hence, we must conclude that the Fourier series converges everywhere, except where it is needed. Let us go one step further and consider a current flowing in a Hertzian dipole. The electric and magnetic field strengths in the far zone are proportionate to the first derivative of the current. If the current is represented by a series expansion, we cannot differentiate term by term to obtain the field strengths, since convergence of a series does not imply convergence of the differentiated series. To obtain some idea about the number of solutions of the wave equation or Maxwell’s equations that cannot be represented with uniform convergence by a Fourier series or transform, let us note that for each solution with uniform convergence there are infinitely many solutions without uniform convergence.

It is worthwhile returning to astronomy at this point. An elliptical orbit
can be represented by a sum of circles without any problems of convergence or differentiability. The concept of epicycles was not wrong—it was only unnecessarily complicated. The simplification of the representation of the orbits of the planets by means of ellipses with the Sun in one focal point led in due course to the theory of gravitation. There is hardly a better example to show the importance of simplicity. Even if a series expansion is used correctly, it may obscure features that a simpler representation would reveal.

A look at the practical side of the convergence problem shows that circuit design is well ahead of theory. The typical on–off type switching functions preferred by semiconductor circuits do not permit an approximation of the transients due to the Gibbs phenomenon, and the transients are the important parts of the switching functions. But this is no problem, since nobody builds pulse generators that contain many amplitude-, frequency-, and phase-stable sinusoidal oscillators in order to produce two-valued pulses according to the Fourier series. On the contrary, it is general practice to synthesize stable sinusoidal oscillations by means of block pulses generated by digital circuits.¹

**Basic Mathematical Concepts**

To see in which way one may profitably generalize our theory of communications based on sine–cosine functions, let us consider Fig. 0-2 again. Block pulses, which were the historically first important system of functions, are shown on the right. The sine–cosine functions plus the constant function used in the Fourier series are shown on the left. One may readily see why we have an extensive theory based on sine–cosine functions, but not one based on block pulses. The block pulses differ by a time shift only. In other words, they contain one free parameter, which we call *delay*. The periodically continued sine–cosine functions contain the parameter delay too, which is called *phase* for these particular functions, but in addition they contain the parameter frequency. In essence, sine–cosine functions of different frequency have a different shape, while the block pulses all have the same shape. For a satisfying, more general theory, one will thus have to look for nonsinusoidal functions that have at least as many parameters as the sinusoidal functions. Since sine–cosine functions are a particular system of orthogonal functions, one may replace them by general systems of orthogonal functions.

The term *orthogonal* is defined as follows: Two functions \( f(j, \theta) \) and \( f(k, \theta) \) with the variable \( \theta \) and the parameters \( j \) and \( k \) are called orthogonal in the interval \(-\frac{1}{2} < \theta < \frac{1}{2}\) if the integral \( \int_{-1/2}^{1/2} f(j, \theta)f(k, \theta) \, d\theta \) is zero for \( j \neq k \). They are called orthogonal and normalized or orthonormal if the integral equals 1 for \( j = k \).

¹ An even better synthesis by means of Walsh functions was reported by Kitai (1975b).
The sine-cosine functions with the normalized time $\theta = t/T$ as variable in Fig. 0-2 are one of the best known systems of orthonormal functions. These functions, including the constant function shown on top, are used in Fourier series expansions.

Walsh functions are another example of a system of orthonormal functions (Walsh, 1923; Fowle, 1905; Osborne, 1918; Pinkert, 1919). A few of them are shown in Fig. 0-2. These functions assume the two values $+1$ and $-1$ only, which makes them comparably simple as the block pulses and leads to simple equipment. On the other hand, the functions have different shapes like the sinusoidal functions, which assures that theory and equipment are not as restricted as for block pulses.

The notation sal($i$, $\theta$) and cal($i$, $\theta$) is used for Walsh functions. The letters $s$ and $c$ allude to sine and cosine functions to which the respective Walsh functions are closely related; the letters $a$ and $l$ are derived from the name Walsh. The functions cal($i$, $\theta$) are even functions like $\sqrt{2} \cos 2\pi i \theta$; the functions sal($i$, $\theta$) are odd functions like $\sqrt{2} \sin 2\pi i \theta$.

In Fig. 0-2 the parameter $i = 1, 2, \ldots$ in $\sqrt{2} \sin 2\pi i \theta$ and $\sqrt{2} \cos 2\pi i \theta$ gives the number of oscillations in the interval $-\frac{1}{2} \leq \theta < \frac{1}{2}$, which is the normalized frequency $i = fT$. One may interpret $i$ as one-half the number of zero crossings per unit time rather than oscillations per unit time (Mann, 1943; Stumpers, 1948; Voelcker, 1966). The zero crossing on the left side, $\theta = -\frac{1}{2}$, but not the one on the right side, $\theta = +\frac{1}{2}$, of the time interval is counted for sine functions.

The parameter $i$ also equals one-half the number of zero crossings in the interval $-\frac{1}{2} \leq \theta < \frac{1}{2}$ for the Walsh functions in Fig. 0-2. In contrast to sine-cosine functions, the sign changes are generally not equidistant. If, unlike in Fig. 0-2, $i$ is not an integer, then it equals one-half the average number of zero crossings per unit time. The term normalized sequence$^1$ has been introduced for $i$, and $\varphi = i/T$ is called the nonnormalized sequence, which is measured in zps:

\[
\text{sequence in zps} = \frac{1}{2} \text{(average number of zero crossings per second)}
\]

The concepts of period of oscillation, $\tau = 1/f$, and wavelength, $\lambda = v/f$, are connected with frequency. Substitution of sequence $\varphi$ for frequency $f$

---

$^1$ The term *sequence* as well as the notation sal($i$, $\theta$), cal($i$, $\theta$), and wal($i$, $\theta$) resulted from joint work by R. Liedl, G. Lochs, F. Pichler, P. Weiss—all then at Innsbruck University in Austria—and the author from 1964 to 1968. The first use of sequence in a publication was by Pichler (1967). The term *sequence theory* developed over the years and was first used in a book title, “Applications of Walsh Functions and Sequence Theory,” by G. Sandy and H. Schreiber in 1974 for a book published by the Institute of Electrical and Electronics Engineers, New York. The following terms have been used for sequence in the scientific literature: *sequentie* (Dutch), *séquence* (French), *Sequenz* (original German term), *sequencia* (Italian), *chastost’* (Russian, in analogy to *chastota*, frequency), *secuencia* (Spanish), *sekvens* (Swedish).
leads to the more general concepts of average period of oscillation, $\tau = 1/\phi$, and average wavelength, $\lambda = v/\phi$:

average period of oscillation = average separation in time of the zero crossings multiplied by 2

average wavelength = average separation in space of the zero crossings, multiplied by 2 ($v$ is the velocity of propagation of a zero crossing$^{1}$)

One of the most important features of sine–cosine functions is that most functions used in communications can be represented by a superposition of them, for which Fourier analysis is the mathematical tool. The transition from time to frequency functions is a result of this analysis. It is often taken so much for granted that one instinctively sees a superposition of sine–cosine functions in the output voltage of a microphone or a teletype transmitter. Actually, the representation of a time function by sine–cosine functions is only one among many possible ones. Complete systems of orthogonal functions generally permit series expansions that correspond to the Fourier series. There are also transforms corresponding to the Fourier transform for many systems of functions. Hence, one may see a superposition of Walsh functions, Legendre polynomials, parabolic cylinder functions, and so forth in the output voltage of a microphone.

The transition from the system of sine–cosine functions to general systems of orthogonal functions brings simplifications, as well as complications, to the mathematical theory of communications. For instance, one may avoid the troublesome fact that any signal occupies an infinite section of the time–frequency domain by substituting a time–function domain. Any time-limited signal containing finite information occupies a finite section of this time–function domain. Most of the simplifications are derived from the fact that periodic sine–cosine functions are inherently unsuitable to describe time–variable systems. The complications, apart from the transitory difficulty of getting used to something new, can usually be traced to the presence of one more parameter than sine–cosine functions have. We may recognize

$^{1}$ The velocity of propagation is an area where the difficulty of using infinitely long, periodic sinusoidal functions in communications shows up clearly. Our intuitive concept of velocity cannot be applied. A phase and a group velocity have been defined, but both can become infinite under certain conditions. Sommerfeld (1914) defined velocities for sinusoidal pulses that are zero outside a finite interval, but these definitions cannot be used for the infinite periodic functions and time-invariant systems. The velocity of propagation of a zero crossing does not cause problems as phase and group velocity do. Let the Walsh functions in Fig. 0-2 be zero outside the interval $-\frac{1}{2} \leq \theta < \frac{1}{2}$ to see that the zero crossings cannot propagate faster than the energy of the signal, that many types of distortions will cause no ambiguity, and that the velocities of beginning and end of the signal are of little or no concern for signals with many zero crossings. Similar comments apply to other functions that are zero outside a finite interval.
this difference by letting sine–cosine and Walsh functions in Fig. 0-2 continue periodically. It is readily verified that no shifting, stretching, or compressing will transform the Walsh functions for \( i = 3 \) into the others, while all sine–cosine functions can be produced from one by these three operations.

**TIME AND SPACE SIGNALS**

The acid test of any theory in science is its practical application. The greatest progress of sequency theory toward practical use was in the area of filters or real-time processors for time and space signals. Filters for time signals have been developed that use the correlation principle and are implemented by multipliers, integrators, adders, and storage circuits. Furthermore, filters using the resonance principle have been implemented with the usual coils and capacitors supplemented by switches.

Much more important are filters for space and space–time signals. The growing interest in PCM television for national satellite links and cable TV creates the market for such filters. The usual theory of filters for time-variable signals cannot be used for spatial filters. Digital computers have been applied very successfully to spatial filtering when fast operation was not required. These computer techniques are discussed in two excellent books by Andrews (1970, 1972). Analog and digital circuits operating in real time have been built with the usual electronic components for the data reduction of PCM television signals first in Japan, and then in England, the U.S., and West Germany.

Of particular interest are the sampling devices based on liquid crystals, which yield inherently transformed signals as required for spatial signal processing. Such devices were developed independently in Japan, using nematic liquid crystals, and in West Germany, using the DAP effect of liquid crystals.

**OPTICAL IMAGES FROM ACOUSTIC WAVES**

One of the best results of sequency theory is a method for the generation of moving images by means of sound waves. This is the first and so far only method that can practically produce moving images of a quality comparable to that of television pictures underwater. We could produce images in the past by means of the lens, by the echo principle as used in radar and sonar, and by holography. Imaging equipment using acoustic lenses underwater would be too bulky for the wavelengths of main interest, which are in the order of 1 cm and longer. Sonar and holography are too slow to produce the 20 to 30 images per second required for the sensation of movement. The spatial electric filters obtained as a result of sequency theory are fast enough to do the job. The filters used for imaging work with sinusoidal sound waves
and could have been built before the introduction of sequency theory into underwater acoustics. However, the direct invention of spatial electric filters for sinusoidal functions was apparently too difficult and they were found via Walsh functions.

The principle of image generation by sound waves and spatial electric filters is as follows. A sound projector insonifies the object plane. Each point in the object plane returns a scattered wave. The sum of all scattered, spherical waves produces the wavefront of the scattered waves in the object plane. This wavefront propagates to the reception plane. There will be a linear relationship between the wavefront in the object plane and the wavefront in the reception plane. If one knows this linear transformation and one can perform the inverse transformation, one regains the wavefront in the object plane. The regenerated wavefront is the image (of the object plane).

For the practical implementation of the inverse transformation one converts the acoustic wavefront into a two-dimensional spatial array of electric voltages. This is done by means of a two-dimensional hydrophone array. The output voltages of the hydrophones are fed to the input terminals of a two-dimensional spatial electric filter. It performs the inverse transformation at the rate of about $10^5$ transforms per second. At the output terminals of the filter one obtains voltages that are proportionate to the amplitudes of the wavefront in the object plane. In other words, one obtains a representation of the acoustic image by electric voltages.

Since only our eyes and not our ears can receive two-dimensional spatial signals, we do not want an acoustic image but rather an optical one. This is readily done by feeding the output voltages of the filter to an electrooptical display such as a TV monitor or an array of light-emitting diodes. The brightness of the diodes then represents the amplitudes of the scattered sound waves at the corresponding points in the object plane.

**Electromagnetic Waves with General Time Variation**

Most sources of electromagnetic waves do not radiate sinusoidal waves. The existence of nonsinusoidal waves has been known since d’Alembert showed in the eighteenth century that the general solution of the wave equation in one space dimension consists of two arbitrary waves $f(x - ct)$ and $g(x + ct)$ traveling from left to right and from right to left. Nevertheless, the theory as well as the equipment design of electromagnetic waves is almost totally dominated by sinusoidal waves. It is sometimes believed that the general solution of d’Alembert can be represented by a superposition of sine–cosine functions, but it was pointed out before that this is incorrect, both mathematically and practically. There were also conjectures that nonsinusoidal waves could not be used because they lacked certain shift-invariant
features of sinusoidal waves. Mathematically correct proofs for these conjectures existed, but they were based on certain assumptions about the topology of space and time that could not be proved experimentally but only with the help of further unprovable assumptions. Such an impasse can only be resolved by experimental demonstration, and this was done. We now have transmitters and receivers for nonsinusoidal waves that show that mobile communication with such waves works practically as well as theoretically.

The most important result of this experimental work over the long run will probably be that it broke the mental hold of the assumptions made about the topology of space and time, on which the unique role of sine–cosine functions rested. In addition, it was learned that the available technology is sufficiently advanced for the use of nonsinusoidal electromagnetic waves, and that there are many applications previously not heard of. The problem of topology is pursued in Chapter 4, while Chapter 3 is devoted to the practical use of such waves in electronics.

The systematic investigation of possible applications does not start with attempts to solve known problems, since the best solvable problems will not even have occurred to someone used to thinking in terms of sinusoidal waves. A much better approach is first to find differences between sinusoidal waves and a particular system of nonsinusoidal waves. Following this path, one finds uses for the particular nonsinusoidal waves, but in addition one is alerted to features of sinusoidal waves that are not shared by other waves. One may then look for waves that differ strongly from sinusoidal waves in a particular feature. Walsh waves were chosen for the general investigation since they lead to readily implementable equipment. In one application—the discrimination between conducting and nonconducting scatterers—theory and experiment are carried somewhat beyond Walsh waves, since Walsh waves too close to the mathematical ideal are not suitable for the effect.

A great number of effects in which sinusoidal and Walsh or other nonsinusoidal waves differ has been found. All these effects are based on a few mathematical differences:

(a) The derivative or the integral of a sinusoidal function is the same sinusoidal function with a time shift and a changed amplitude. In other words, the shape of the function is not changed. The shape of Walsh functions, on the other hand, is changed by differentiation or integration.

(b) The sum of sinusoidal functions with the same frequency but various amplitudes and phases is a sine function with the same frequency. Other waves sum differently.

(c) The Doppler effect appears in two ways. A single Walsh wave shows the Doppler effect stronger than a single sinusoidal wave. An orthogonal set of Walsh waves, on the other hand, is less affected than an orthogonal set of
sinusoidal waves. A relative velocity $|v/c| > \frac{3}{5}$ is required to transform one Walsh wave of the set into another, while no such minimum velocity exists for sinusoidal waves.

(d) An amplitude reversal of a sine wave is equivalent to a time shift. No such equivalence exists for waves that are not polarity symmetric.

Let us note that the large number of effects for nonsinusoidal waves was not found by accident. Electromagnetic waves are represented by vectors with the four variables $x$, $y$, $z$, and $t$. At the other extreme are voice signals, which are represented by a scalar with the one variable $t$. More degrees of freedom provide a better chance to find something new, and one should devote one’s effort to the field with the best chance of success.

Many of the results derived for nonsinusoidal waves can be explained and understood as being caused by the great frequency bandwidth of these signals. However, one should be cautious not to carry this interpretation too far. For instance, the discrimination in radar of an airplane made of aluminum from raindrops or from the surface of the Earth by means of non-polarity-symmetric waves is not an automatic consequence of using wideband signals. On the quantitative level one must, of course, be even more careful, since all signals have a beginning and an end, and thus occupy an infinite frequency bandwidth.

**Concepts of Communications Applied to Physics**

If nonsinusoidal electromagnetic waves can be generated and received in the radio range, one is forced to consider nonsinusoidal light waves. Spectrometers decompose light into sinusoidal waves, but this does not mean that light cannot be decomposed into nonsinusoidal waves.¹ A spectrometer

¹ It was well known to the physicists of the nineteenth century that the decomposition of light into sinusoidal functions was only one out of many possible ones: “we are at liberty, whenever it is convenient, to present white light by superposing a number of homogeneous vibrations having periods which lie very close together. But we are equally at liberty to assume any other representation so long as its resolution by Fourier’s theorem gives us a distribution of intensity equal to that of the observed one” (Schuster, 1904). “… on peut regarder la lumière blanche comme formée par une suite d’impulsions tout à fait irrégulières, ou de vibrations sans cesse troublées, analogues au mouvement de trépidation qui, pour quelques physicians, constitue le mouvement calorifique” (Gouy, 1886). Rayleigh points out the possibility of non-Fourier decomposition for sound (1894, Vol. 1, pp. 24, 25) but writes as follows about light (1889): “There is nothing arbitrary in the use of a circular function to represent the waves. As a general rule this is the only kind of wave which can be propagated without a change of form; and, even in the exceptional cases where the velocity is independent of wave-length, no generality is really lost by this process, because in accordance with Fourier’s theorem any kind of periodic wave may be regarded as composed of a series of such [sinusoidal waves], with wavelengths in harmonic progression.” Huygens (1690) never mentions sinusoidal waves in his theory of light. Sinusoidal functions did not become preeminent until Fourier’s book was published in 1822.
for the decomposition into Walsh waves has been found. It cannot be built at present, since it requires shutters with operating times of about $10^{-15}$ sec, while the fastest shutters we have, using the Kerr effect, require an operating time upward of $10^{-10}$ sec.

Having found a spectrometer that decomposes light into nonsinusoidal functions, even though it works on paper only, makes one question the fundamental role of sine–cosine functions in physics. The hard core for the role of sinusoidal functions is their distinction by the so-called real-number topology of space–time. This concept of space–time is generally accepted, but it is in complete contradiction to the results of information theory. Let us elaborate this point to recognize the contradiction, learn its cause, and see where we might look for an experimental decision.

Consider a time interval $0 \leq t < T$, where $T$ is arbitrarily large but finite. This is the longest time interval of interest in physics, since no one has had experience with infinite time intervals or ever expects to have such experience. According to the accepted concept of time, there are as many time points in this interval as there are real numbers between the number 0 and the number $T$. For any relationship between time points there is an exactly equal relationship between real numbers. For instance, $t + t_0$ is a time point later than $t$, and the number $t + t_0$ is larger than the number $t$. Let some event or sequence of events be described by a time function $F(t)$. If the same events had happened with a delay time $\Delta t$ they would be described by $F(t - \Delta t)$.

Let us map the time interval $0 \leq t < T$ by means of the exponential function $\exp(i2\pi t/T)$ onto the unit circle. There is a one-to-one relationship between the points in the interval $0 \leq t < T$ and the points on the circle. We could have used a hexagon or some other curve for the mapping and still have obtained a one-to-one relationship. However, the effect of a delay time $\Delta t$ would be different. A delay $\Delta t$ leaves the exponential function $\exp[i2\pi(t - \Delta t)/T] = \exp(-i2\pi \Delta t/T) \exp(i2\pi t/T)$ unchanged except for a factor. A mapping on any other curve but the circle would not preserve the shift invariance of the processes $F(t)$ and $F(t - \Delta t)$.

If one is convinced by this argument that time distinguishes the circle and the exponential function, one will readily accept sine–cosine functions as distinguished, since they are merely a way of writing the exponential function with real and imaginary components or in a cartesian coordinate

---

1 Any point on a straight line or a circle has the same relationships to the points in its neighborhood, if we exclude the beginning and end of the line. The points at the corners of a hexagon, on the other hand, have a different relationship to their neighbors than the points on the straight sections between the corners. The mathematician says that the exponential function is the characteristic group of the topologic group of the real numbers.
system. Furthermore, one will readily accept that this argument still holds if the time variable $t$ is replaced by one space variable $x$, and the time interval of duration $T$ by a space interval of length $X$.

The obvious trouble with our argument is that it must not only hold for finite delay times $\Delta t$ but also for infinitely short or—more precisely—infinitesimal delay times $dt$. Otherwise one could substitute a regular polygon for the circle with $T/\Delta T$, or perhaps $2T/\Delta T$, equally long sides. The measurement of infinitesimal intervals $dt$ is certainly no more part of our current or foreseeable experience than the measurement of infinite intervals $T \rightarrow \infty$. Let us derive a quantitative statement of this fact.

In order to measure the location of a time point $t_0$ in the interval $0 \leq t < T$ one may divide the interval into a left and a right half. If $t_0$ is located in the right half we write 1, and if it is located in the left half we write 0. The two intervals $0 \leq t < T/2$ and $T/2 < t < T$ are again divided into right and left halves. The location of the point $t_0$ in one of the four quarters $0 \leq t < T/4$, $T/4 \leq t < T/2$, $T/2 \leq t < 3T/4$, $3T/4 \leq t < T$, is then denoted by the binary numbers 00, 01, 10, 11. Further subdivisions lead to greater accuracy of measurement or to a better knowledge of the location of point $t_0$, and the binary number describing the location gets more and more digits.\(^1\) Each binary digit contains one bit of information. Hence, by performing the time measurement as described, we can say how much information in bits we have about the location of the point $t_0$. Let the accuracy of measurement increase beyond all bounds. The binary number describing the location of $t_0$ approaches denumerably many digits and the information becomes denumerably infinite. This is clearly beyond reality. There is no way to store, process, or transmit denumerably many binary digits. We do not have to speculate about the accuracy of various methods of measurement. They may be accurate beyond comprehension, but we will never use more than a finite amount of information.\(^2\)

We have, however, not yet reached the limit. When the number of subdivisions has become denumerably infinite, it can keep on growing and become nondenumerably infinite. We are now two abstractions away from reality but we have found the number of infinitesimal distances $dt$ that cover the interval $0 \leq t < T$.

---

\(^1\) This type of measurement based on yes–no decisions is part of information theory in communications; in physics it is known under names like propositional calculus (Jauch, 1968; Misner et al., 1973). The importance of a finite number of yes–no decisions is familiar to those using data processing equipment, but less so to the pure theorist [see, e.g., Jauch (1968, p. 73, third paragraph, or p. 97, last paragraph)].

\(^2\) To illustrate the importance of the finite amount of transmittable information, let us observe that the finite velocity of the transmission of information—or the finite propagation velocity of signals—is the basis of the special theory of relativity.
Having shown that the concept of real-number space–time is in gross contradiction to our ability of information processing, we must next show where it came from. The use of the differential $dt$ is the giveaway. The concept of the real-number topology\(^1\) of space–time comes from differential calculus. One cannot use differential calculus without assuming this topology. Of course, this proves nothing more about physical space–time than that the mathematical model is applicable to it whenever the results derived from the model correspond with experimental results. Differential calculus is a convenient mathematical method, but we must not permit it to define the topology of space–time any more than we permit Euclidean geometry to define its metric. Putting it somewhat more simply, planar geometry is convenient, but this does not make our Earth flat.

The mention of Euclidean geometry clearly indicates how to proceed. The hold on thinking by this geometry was broken through the development of two other models, the spherical and the pseudospherical geometry.\(^2\) They paved the way to Riemann geometries. We will then investigate a special model of topology of space–time. From the way the measurement of the point $t_0$ in the interval $0 \leq t < T$ was discussed, one would expect this special model to favor the Walsh functions. However, luck will be on our side. It turns out that a whole class of topologies yields the same result if we consider the movement of one point or one particle only. The real-number topology does not belong to this class, and this opens a way toward results that differ from the ones derived with the help of real-number topology. Different results based on the distinction between finite differences $\Delta t$, $\Delta x$, and differentials $dt$, $dx$ are most likely to be found where time differences and distances are smallest, that is, in nuclear physics. It will be shown that the non-real-number topologies lead to deviations for a relativistic particle in a Coulomb field at distances of some Compton wavelengths from the center of the field.

A GUIDE TO READING

The scope of this book ranges from rather abstract mathematics to fairly detailed equipment design. This mixing of theory and application was done since only experimental and applied work can decide whether a theoretical development is a useful generalization or a useless abstraction. Most

\(^1\) The mathematician calls this the “usual topology of the real numbers,” since one can define more than one topology on the real numbers.

\(^2\) English translations of the publications of Bolyai (1832) and Lobachevski (1840) may be found in a reprint of a book by Bonola (1912).
scientists are interested in either theoretical or experimental work, and the following comments may help to find the right sections.

Most of Chapter 1 should be read by all, even though some sections will do no more than familiarize the reader with the notation. Excepted are Sections 1.1.8, 1.2.4–1.2.6, and 1.4.1–1.4.4, which deal with data processing and are not required for the rest of the book; Sections 1.3.2 and 1.3.3 are only needed by those who want to read Chapter 4.

Chapters 2–4 are essentially independent of each other with the following exceptions: Section 2.1.1 is of general interest; Sections 3.7.1–3.7.6 are also applicable to underwater acoustics, particularly in connection with Sections 2.4.1–2.4.4 on acoustic imaging; Sections 4.1.1–4.1.2 are of general interest to illustrate the meaning of time-variable devices.