1951

# Square wave modulation

Richard Henry Duncan

# SQUARE WAVE MODULATION

BY

RICHARD H. DUNCAN

-----



A

THESIS

submitted to the faculty of the

SCHOOL OF MINES AND METALLURGY OF THE UNIVERSITY OF MISSOURI

in partial fulfillment of the work required for the

Degree of

MASTER OF SCIENCE, PHYSICS MAJOR

Rolla, Missouri

1951

-----

Approved by - Louis H. Lund
Professor of Physics

# TABLE OF CONTENTS

## ACKNOWLEDGMENT

The author wishes to acknowledge the assistance and interest of Dr. Louis H. Lund of the Physics Department and Dr. Janos Zaborszky of the Electrical Engineering Department of the Missouri School of Mines and Metallurgy.

# LIST OF ILLUSTRATIONS

# INTRODUCTION

Modulation of a radio frequency carrier can be defined as the variation of some wave parameter such as amplitude, frequency, or phase, in accordance with intelligence to be transmitted. An unmodulated carrier is simply a continuous sine wave and contains no information other than the fact that the transmitter has been turned on. The spectrum of an unmodulated signal consists of a single frequency. Any type of modulation is characterized by the appearance of additional frequencies, known as sidebands. One of the fundamental problems in radio engineering is to reduce the number of sidebands to a minimum and still adequately represent the information to be transmitted. The large number of radio facilities required by modern civilization can be accommodated in the radio spectrum only if each facility uses bandwidth as sparingly as possible. Reduction of bandwidth inevitably results in lower quality service, so that final designs are compromises between conflicting considerations. The purposes of any modulation study are to provide an understanding of the modulation process and a basis for reducing bandwidth. Obviously, the engineer cannot efficiently take measures to reduce bandwidth until he knows the ideal bandwidth requirement for a system.

A study of square wave modulation is of importance in the field of radio telegraphy; telegraphic signals are characterized by sudden transitions between two signal conditions in both on-off keying and frequency shift keying. It is the purpose of this paper to study the signal spectra resulting when a carrier amplitude is modulated by a square wave. These two types of modulation correspond to on-off keying and frequency shift keying, respectively.

## REVIEW OF LITERATURE

There is already considerable literature dealing with quality of service versus bandwidth in both radio broadcasting and telegraphy. A representative list of such articles is given in the Bibliography of this paper. The results published in these papers have been, for the most part, obtained through experience with actual telegraphic facilities. Although they are of considerable importance in telegraphic art, only one of these publications deals directly with the problem of this paper, the theoretical spectrum resulting from square wave modulation. That article was written by the Dutch engineer and mathematician, Balth Van der Pol.[1] The work of Van der Pol would be rigorous in the case of pure frequency modulation without phase shift of the radio frequency carrier. It is shown in this paper that phase shift does occur, and in such a way as to insure continuity of the resulting signal. When the condition that the final signal be continuous is considered, it appears that more than one type of signal spectrum is possible, including that found by Van der Pol.

(1) Van der Pol, Balth, Frequency Modulation, Proc. I.R.E., Vol. 18, pp. 1194-1205, July 1930.

In addition to the articles listed in the Bibliography, free use has been made of the fundamentals of modulation theory as usually found in texts on radio engineering. References have been given where applicable. These references are for convenience and are not meant to cite original papers in the field.

## DISCUSSION

### 1. Spectrum of a Carrier On-Off Signal

The first signal to be considered consists of an infinite train of sinusoidal pulses, the durations of the "signal on" and "signal off" conditions being the same. This condition is convenient to handle mathematically and is, to a close degree of approximation, representative of the test signal ...RYRYRY... used in teletype practice.[2]

(2) Watson, E. F., Fundamentals of Teletypewriters Used in the Bell System, Bell System Tech. Journal, p. 620, October 1938.

The mark and space arrangement of this combination of letters in the teletype code is such as to require the reversal of the relays in the system at the end of almost every unit of time. Naturally, such a signal constitutes the most stringent test of the system. Furthermore, it requires the most frequent starting and stopping of the carrier (or shifting of the carrier when frequency shift keying is used). Any other combination of letters would insure the existence of the same signal conditions for more than one unit of time; it is well known from Fourier analysis that a sinusoidal pulse approaches monochromaticity as the length of the pulse is increased. Hence, it seems reasonable that the signal corresponding to square wave keying will give the widest spectrum to be encountered in a teletype system.

The signal to be analysed is shown in Figure 1. A few cycles of the radio frequency oscillation are shown. It is impracticable to draw a high frequency oscillation to scale, and it should be understood that there are a large number of cycles in each pulse. A typical case might be a carrier frequency of three megacycles per second and a pulse duration of 1/30 second, giving 100,000 cycles of radio frequency in each pulse.

Fourier analysis of the signal is possible only if the signal is periodic. The mathematical condition of periodicity for the signal can be stated as $f(t_1) = f(t_1 \pm 2T)$ where $t_1$ can have any value, and 2T is the fundamental period of the keying wave. (The possibility of periodicity based on several keying intervals will not be considered in this paper; the condition of periodicity will be restricted to the one given.) Physically, this means that each pulse must begin with the same phase if Fourier analysis is to be applicable. Keeping this necessary condition in mind, the signal can be written as a function of time as follows:

$$(1)\quad \begin{aligned} f(t) &= A \sin(\omega_1 t + \alpha_m) \qquad & (2n+1)T > t > 2nT \\ f(t) &= 0 & 2nT > t > (2n+1)T \end{aligned}$$

The phase angles $\alpha_m$ in the above expression must be free to take on whatever values are needed in order to satisfy the condition of periodicity. (This is perfectly analogous to representing the parts of a saw tooth wave by $f(t) = \pm a_t$

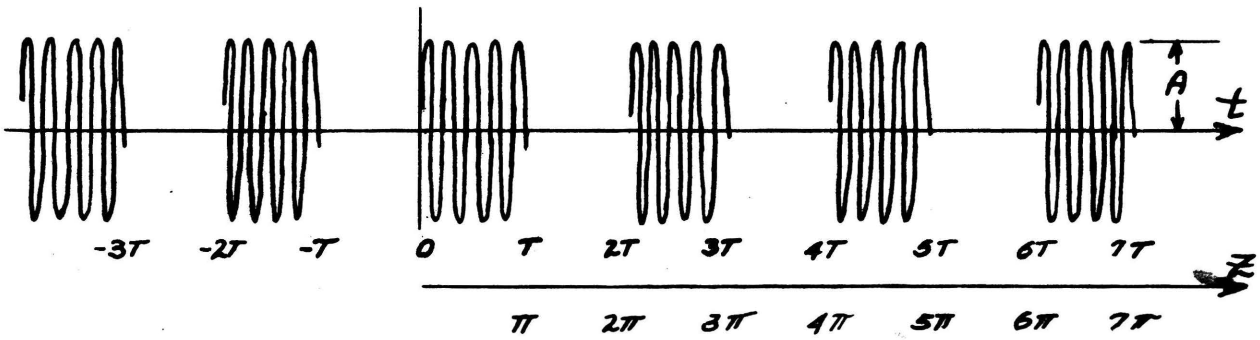


FIG 1. ON-OFF KEYED WAVE.

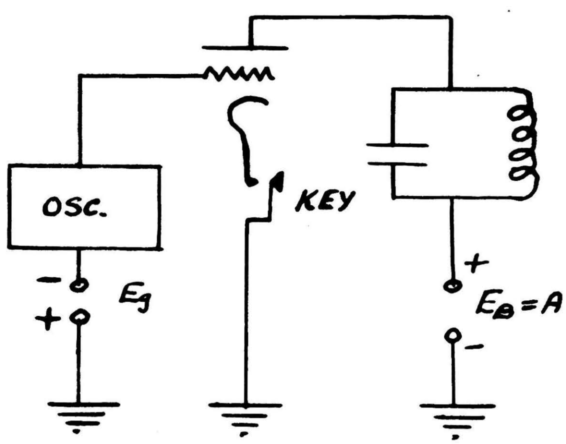


FIG 2. KEYED TRANSMITTER.

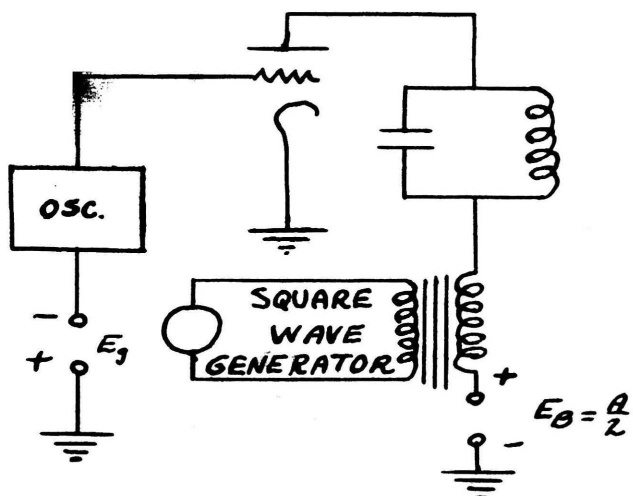


FIG 3. MODULATED TRANSMITTER.

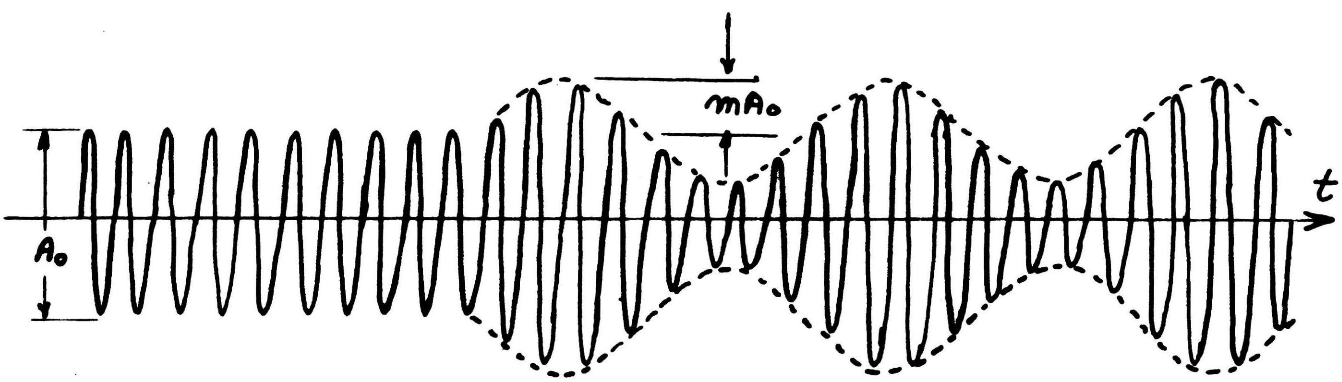


FIG. 4. AMPLITUDE MODULATED WAVE.

$\pm b_n$; the constants $b_n$ are necessarily different for each part of the saw tooth outline.) An interesting result is obtained when there are an integer number of cycles of radio frequency in each pulse; such a signal could be presumed to be generated by keying the ideal radio frequency amplifier shown in Figure 2. The amplifier is fed by an oscillator which is continually on; the amplifier itself is turned on and off by means of the key. Absence of starting and stopping transients is assumed. If the oscillator is ideal, that is, perfect frequency and phase stability are presumed, each pulse of radio frequency in the output will begin with the same phase in the case under consideration. Mathematically speaking, if the carrier frequency is an integer harmonic of the keying frequency the $\alpha_n$ of (1) will all be equal.

Before beginning the Fourier analysis for this special case, it is necessary to change the variable t which ranges from 0 to 2T during a fundamental period to one which ranges from 0 to $2\pi$ during the same period. This change is shown by the Z axis in Figure 1. It can be seen that $t = \frac{T}{\pi} z$ . We can also write $\omega_1 t$ in terms of z by noting that if $N_1$ is the number of cycles of radio frequency in time T, then $N_1 = f_1 T$. Hence, $\omega_1 t = \omega_1 \frac{T}{\pi} z$ . But since $\omega_1 = 2\pi f_1$ and $f_1 = \frac{N_1}{T}$ , the equations (1) can be rewritten as:

$$(2) \quad f(z) = A \sin(2N_1 z + \alpha) \qquad \pi > z > 0$$

$$f(z) = 0 \qquad 2\pi > z > \pi$$

The Fourier coefficients are then given by

$$(3)\quad a_n = \frac{1}{\pi}\int_0^{\pi} A\sin(2N_1 z+\alpha)\cos nz\,dz + \frac{1}{\pi}\int_{\pi}^{2\pi}(0)\cos nz\,dz$$

$$(4)\quad b_n = \frac{1}{\pi}\int_0^{\pi} A(\sin 2N_1 z+\alpha)\sin nz\,dz + \frac{1}{\pi}\int_{\pi}^{2\pi}(0)\sin nz\,dz$$

Reducing (3) and (4) by the trigonometric identity for the sine of the sum of two angles, they become

$$(5)\quad a_n = \frac{A\cos\alpha}{\pi}\int_0^{\pi}\sin 2N_1 z\cos nz\,dz + \frac{A\sin\alpha}{\pi}\int_0^{\pi}\cos 2N_1 z\cos nz\,dz$$

$$(6)\quad b_n = \frac{A\cos\alpha}{\pi}\int_0^{\pi}\sin 2N_1 z\sin nz\,dz + \frac{A\sin\alpha}{\pi}\int_0^{\pi}\cos 2N_1 z\sin nz\,dz$$

The right hand integral of (5) is 0, except when $n = 2N_1$. Noting this fact and integrating

$$(7)\quad a_n = \left\{\frac{A\cos\alpha}{\pi}\left[-\frac{\cos(2N_1-n)z}{2(2N_1-n)} - \frac{\cos(2N_1+n)z}{2(2N_1+n)}\right]_0^{\pi}\right\} + \frac{A\sin\alpha}{\pi}\left[\frac{z}{2}\right]_0^{\pi}$$

$$(\text{If } n = 2N_1)$$

The bracketed part of $a_n$ will be handled first, calling it $a_n'$ and substituting limits,

$$(8)\quad a_n' = \frac{A\cos\alpha}{\pi}\left[-\frac{\cos(2N_1-n)\pi}{2(2N_1-n)} - \frac{\cos(2N_1+n)\pi}{2(2N_1+n)} + \frac{1}{2(2N_1-n)} + \frac{1}{2(2N_1+n)}\right]$$

When n is even, the cosine terms in the numerators of the first two terms in the bracket will be 1, since $2N_1$ is certainly even $(2N_1 \pm n)\pi$ will be an even number of radians. Hence, the entire bracket will be zero when n is even, unless it is equal to $2N_1$ in which case $a_n$ will have to be evaluated separately.

For the special case when $n = 2N_1$

(9) $$a_n' = \frac{A \cos\alpha}{\pi} \int_0^{\pi} A \sin 2N_1 z \cos 2N_1 z \, dz = 0$$

This result is obtained by noting the trigonometric identity $\sin x \cos x = \frac{1}{2} \sin 2x$, which when integrated from 0 to $\pi$ yields 0.

Returning to the expression for $a_n$, it is seen that when n is odd, the cosine terms in the numerators of the first two terms will be -1 since $(2N_1 \pm n)\pi$ will always be an odd number of $\pi$ radians. Reducing (7) and summarizing, the results thus far obtained

(10) $$a_n = \frac{A \cos\alpha}{\pi} \left[ \frac{4N_1}{4N_1^2 - n^2} \right] \qquad n \text{ odd.}$$

$$a_n = \frac{A \sin\alpha}{2} \qquad n = 2N_1$$

The first integral in (6) will vanish except when $n = 2N_1$ in which case it will yield $b_n = \frac{A \cos\alpha}{2}$. The second integral in (6) is like the first in (5), except that the roles of $2N_1$ and n are interchanged. This interchange will not affect the evaluation of $\cos(2N_1 \pm n)\pi$, but it will reverse the signs in the denominators of the terms in (8). Hence, for $b_n$ we write

(11) $$b_n = -\frac{A \sin\alpha}{\pi} \left[ \frac{4N_1}{4N_1^2 - n^2} \right] \qquad n \text{ odd}$$

$$b_n = \frac{A \cos\alpha}{\pi} \qquad n = 2N_1$$

Letting the quantity $\frac{4N_1}{4N_1^2 - n^2} = B$ , the Fourier expansion of the signal can now be written as

$$(12)\quad f(z) = \frac{A\sin\alpha}{2}\cos 2N_1 z + \frac{A\cos\alpha}{\pi}\sum_{n=1,3,5,\cdots}^{\infty} B\cos nz + \frac{A\cos\alpha}{2}\sin 2N_1 z - \frac{A\sin\alpha}{\pi}\sum_{n=1,3,5,\cdots}^{\infty} B\sin nz$$

The Fourier expansion (12) can be further simplified. The $\sin 2N_1 z$ and $\cos 2N_1 z$ terms are of the same frequency and can be combined.

$$(13)\quad \frac{A\sin\alpha}{2}(\cos 2N_1 z) + \frac{A\cos\alpha}{2}(\sin 2N_1 z) = C\sin(2N_1 z + \psi)$$

Expanding the right hand side of (13) by the trigonometric identity for the sine of the sum of two angles, equating coefficients of like terms and solving for $\psi$ and C in the usual manner will give $C = A/2$ and $\psi = \alpha$ . A similar combination of corresponding terms occurring under the summation signs will give terms of the type

$\frac{A}{\pi}\frac{4N_1}{4N_1^2 - n^2}\cos(nz + \alpha)$ where B has been replaced by its value in terms of $2N_1$ and n. The Fourier expansion can now be rewritten as

$$(14)\quad f(z) = \frac{A}{2}\sin(2N_1 z + \alpha) + \sum_{n=1,3,5,\cdots}^{\infty}\frac{A}{\pi}\frac{4N_1^2}{4N_1^2 - n^2}\cos(nz + \alpha)$$

It is convenient at this time to change the variable back to t. Remembering that $z = \frac{\pi}{T}t$ and multiplying numerator and denominator by $2f_1$, $z = \frac{\omega_1 t}{2N_1}$. The Fourier expansion can now be written in terms of t.

$$(15) \quad f(t) = \frac{A}{2}\sin(\omega_1 t + \alpha) + \sum_{n=1,3,5,\ldots}^{\infty} \frac{A}{\pi}\frac{4N_1}{4N_1^2 - n^2}\cos\left(\frac{n\omega_1 t}{2N_1} + \alpha\right)$$

It has already been noted than $N_1$ is a very large number, the number of cycles of radio frequency occurring during the period of an audio frequency. For this reason, it will be convenient to rearrange the counting starting with $2N_1$ and counting ahead and back from $2N_1$; thereby avoiding computations using the squares of large numbers. Besides the added convenience, the rearrangement of the counting will reveal a characteristic of the signal which is in itself interesting. We begin by letting $n = 2N_1 \pm K$ where k is odd, then n will be odd as required in the above summations. Substituting for the kth frequency of the series (15)

$$(16) \quad \frac{(2N_1 \pm K)\omega_1 t}{2N_1} = \left(\frac{2N_1\omega_1}{2N_1} \pm \frac{K\omega_1}{2N_1}\right)t = \left(\omega_1 \pm \frac{2\pi f_1 K}{2f_1 T}\right)t$$

$$= (\omega_1 \pm K\omega_0)t \quad , \quad f_0 = \frac{1}{2T}$$

Furthermore, the coefficients can be simplified as follows:

$$(17)\quad \frac{A}{\pi}\,\frac{4N_1}{4N_1^2-n^2} = \frac{A}{\pi}\left[\frac{4N_1}{4N_1^2-(2N_1\pm k)^2}\right] = \frac{A\ 4N_1}{\pi[\pm 4N_1 k + k^2]}$$

$$= \frac{A}{\pi k}\,\frac{4N_1}{(\pm 4N_1 + k)}$$

An approximation can be introduced at this point; since $N_1$ is a very large number, $\frac{4N_1}{\pm 4N_1 + k}$ will be very nearly equal numerically to 1 for a large range of values of k beginning with 1. It should also be noted that the magnitudes of the coefficients will be negligibly small due to the 1/k factor long before this approximation will cease to be valid. Hence, the approximation will cause no error in the terms of the series which are retained in any practical computation. Letting k range from $2N_1 - 1$ to infinity so that n will range from 1 to infinity as required, the expansion can now be written

$$(18)\quad f(t) = \frac{A}{2}\sin(\omega_1 t + \alpha) + \sum_{k=1,3,5,\dots}^{2N_1-1} \frac{A}{\pi k}\cos\left[2\pi(f_1 - k f_0)t + \alpha\right]$$

$$- \sum_{k=1,3,5,\dots}^{\infty} \frac{A}{\pi k}\cos\left[2\pi(f_1 + k f_0)t + \alpha\right]$$

The final expression for the spectrum of the signal in Figure 1 will now be written with the limits on the summations omitted; it is to be understood that the summations will be run from 1 to whatever value is desired for computational accuracy, with k odd.

$$(19) \quad f(t) = \frac{A}{2} \sin(\omega_1 t + \alpha) + \sum \frac{A}{\pi k} \cos[(\omega_1 - k\omega_0)t + \alpha] - \sum \frac{A}{\pi k} \cos[(\omega_1 + k\omega_0)t + \alpha]$$

Inspection of (19) will reveal that it is identical to the result that would have been obtained by considering the signal to be composed of a carrier of amplitude A/2, amplitude modulated by a square wave of amplitude A/2. The treatment of amplitude modulation by a single sine wave tone will be reproduced here in order to lend continuity to the discussion. A carrier amplitude modulated by a single tone is depicted in Figure 4. The carrier amplitude is caused to vary by an amount $\pm\, mA_0$ where m is known as the percentage modulation. Mathematically, the signal can be expressed as

$$(20) \quad f(t) = A_0 [1 + m \sin \omega_0 t] \sin(\omega_1 t + \alpha)$$

Subsequent multiplication and rearrangement of terms by means of a trigonometric identity will yield

$$(21) \quad f(t) = A_0 \sin(\omega_1 t + \alpha) + \frac{mA_0}{2} \cos[(\omega_1 - \omega_0)t + \alpha] - \frac{mA_0}{2} \cos[(\omega_1 + \omega_0) t + \alpha]$$

In order to derive (19) from amplitude modulation theory, a more general type of expression for amplitude modulation is needed. When a carrier is modulated by an arbitrary function of time, the resulting signal can be written as (3)

(3) Goldman, Stanford, Frequency Analysis, Modulation and Noise, McGraw-Hill Book Co., pp. 141-179, 1948.

(22) $$f(t) = A_0 \left[1 + m\, g(t)\right] \sin\left[\omega_1 t + \alpha\right]$$

In the above expression, $mA_0 g(t)$ is the modulating function of time. For the signal under consideration here, $g(t) = \sum_{K=1,3,5,\cdots}^{\infty} \frac{2A}{\pi K} \sin K\omega_0 t$, the Fourier expansion of a square wave, the percentage of modulation is 1.00 if we consider the original carrier amplitude to be A/2. Substitution into (22) gives

(23) $$f(t) = \frac{A}{2}\left[1 + 1.00 \sum \frac{2}{\pi K} \sin K\omega_0 t\right] \sin(\omega_1 t + \alpha)$$

Multiplication and rearrangement of the terms by means of a trigonometric identity gives (19) identically, thus justifying the statement that the signal spectrum can be derived from either Fourier analysis or amplitude modulation theory.

This result is not at all surprising; in the ideal case, modulation and keying amount to the same thing physically. Consider the idealized transmitter which has been referred to and is shown in Figure 2. (If one is to analyse idealized signals, he must be permitted to conceive of ideal transmitters.) The effect of keying is to vary the plate voltage of the amplifier from 0 to A; a supply of voltage A being assumed. The oscillator is on continuously, and its output will be amplified anytime plate voltage is applied to the tube; and if the grid driving voltage is high enough, an oscillation of amplitude A will occur in the plate tuned circuit. Next, consider the idealized transmitter of Figure 3 which has a plate supply voltage of A/2 and is modulated through a transformer of perfectly linear frequency response by a square wave of amplitude A/2. In this case, the plate voltage will vary from 0 on the negative lobes of the square wave to A on the positive lobes. The effect of modulation in this case is then to cause a plate voltage variation from 0 to A. This is the same result which was produced by keying the transmitter of Figure 2. In short, the two operations, modulation and keying produced the same signal.

So far, it appears that the obvious has been proved. Actually, a highly special case, namely, that of a carrier frequency which is an integer harmonic of the keying frequency was assumed. However, three separate cases can occur. They will now be discussed in order.

Case 1. The carrier frequency is an integer harmonic of the keying frequency.

This is the case which was just analysed by two independent methods. If the carrier frequency is assumed to be an integer harmonic of the keying frequency, each pulse will quite naturally and of its own accord start with the same phase. It will be recalled that the oscillators in the idealized transmitters were considered to be on continuously. Thus, in this case, the mathematical condition of periodicity is satisfied automatically. In the preceding analysis, two independent methods yielded the same result for the signal spectrum, except for an approximation of little practical importance which was introduced in deriving the coefficients of the components of the spectrum by Fourier analysis. The author has not yet been able to resolve the question of why the two analyses do not yield the same results identically without the approximation. While this anomaly has very little practical importance, it is of mathematical interest. It might be related to the Gibbs phenomena which occurs at discontinuities in functions treated by Fourier analysis.

Case 2. The carrier frequency is not an integer harmonic of the keying frequency; the carrier occurs in each pulse with its natural phase.

If the carrier frequency is not an integer harmonic of the keying frequency, and the oscillators of either of the two ideal transmitters are still assumed to operate continuously, the pulses will each start with different phase. Hence, the condition of periodicity will not be fulfilled, and a Fourier series representation based on a single keying interval will not be applicable. However, the carrier component of the spectrum will have constant phase as required by amplitude modulation theory, and the signal spectrum will be given by (19) as before.

Case 3. The carrier frequency is not an integer harmonic of the keying frequency, and each pulse is forced to start with the same phase.

This case is very much like Case 2, except that the condition of periodicity is obtained by adjusting the phase of the oscillators at the beginning of each pulse. Direct Fourier analysis will be applicable; since the carrier component will not have constant phase, modulation theory cannot be applied. As a matter of fact, the carrier itself will not appear in the spectrum since only those frequencies which are harmonics of the fundamental period of integration appear in a Fourier spectrum. If the analysis were carried out, one would expect both even and odd harmonics of the fundamental in the spectrum expansion. Only odd harmonics were found in Case 1 by Fourier analysis and modulation theory. There is no contradiction here; modulation theory is simply not applicable.

Of the three cases discussed, the first two are of the most practical importance. Generation of the type of signal discussed in Case 3 by either of the ideal transmitters would entail the addition of a perfect phase control. The process of phasing and keying a wave is certainly different from that of simply keying the wave, being a combination of phase and amplitude modulation. In any practical case, amplitude modulation theory is simple to apply, and it has been shown to give the correct result. However, it is not meant to discredit the signals of Case 3. All of the ideas presented so far will be used to advantage in the next section in the treatment of frequency shift keying.

2. Spectrum of a Signal Employing Frequency Shift Keying

A frequency shift signal is shown in Figure 5. The common practice of referring to such a signal as an FS signal will be employed here. Once again, a square wave has been taken as the keying wave. The mark condition is denoted by one frequency, the space condition by another. Such signals are derived physically by injecting the keying wave into a reactance tube circuit, changes in tuned circuit reactance being caused by changes in reactance tube current. Such circuits are easily found in the literature and need not be fully discussed here.(4) Physically, then the signal of Figure 5 is the result of frequency modulating a carrier with a square wave.

(4) Terman, Frederick E., Radio Engineering, McGraw-Hill Book Co., p. 493, 1947.

Just as in the case of start-stop keying, it is necessary to consider the phase with which each pulse of radio frequency begins before making a Fourier analysis. One condition should be evident; the signal should be continuous at transition points. A simple mechanical analogy will make this clear. Consider the mechanical oscillator

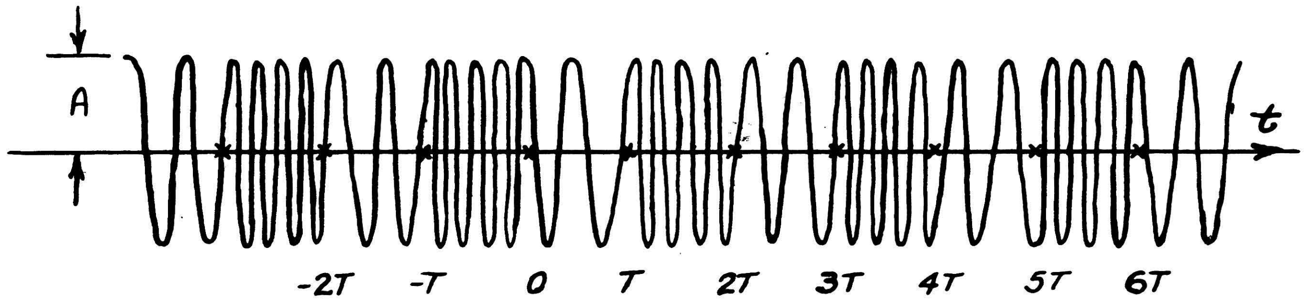


FIG. 5. FS KEYED WAVE.

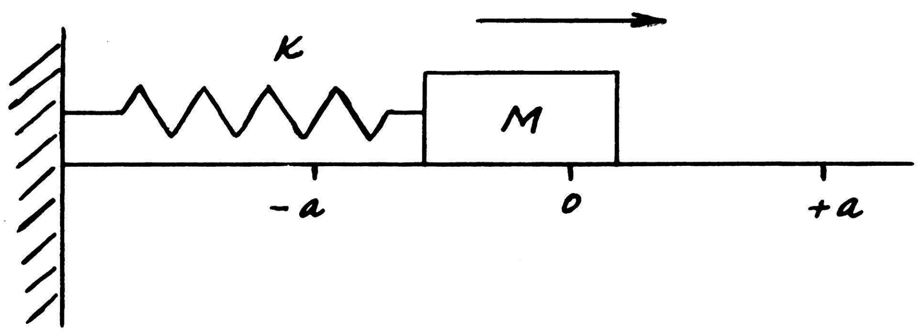


FIG 6. MECHANICAL OSCILLATOR.

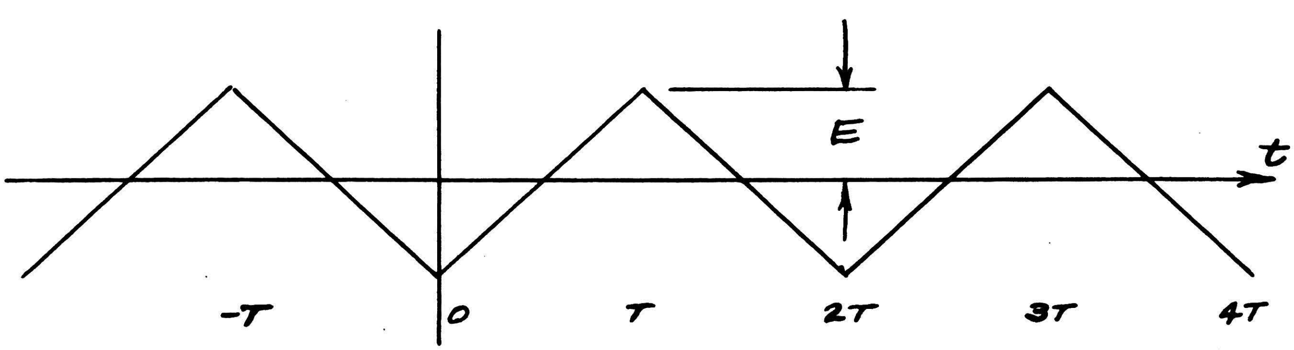


FIG 7. TRIANGULAR WAVE.

of Figure 6. The mass M oscillates about the point O between the limits $\pm$ a, and at the instant under consideration is moving toward the right. Now, consider a sudden change in one of the circuit parameters, either the mass M or the spring constant k. The mass will continue to move to the right, but the oscillation will continue with a new frequency determined by the new circuit parameters. Intuitively, one would require continuity of motion in any mathematical description of the problem. Since transients are being neglected in this treatment, continuity would imply that the oscillatory functions representing the motion in two successive intervals be equal at the transition point between the intervals. Consideration of transients would require that the functions and their derivatives be equal at the transition point.

Analogously, continuity of the functions representing the FS signal will be required. It will be shown that such continuity follows quite naturally from frequency modulation theory. The nature of the FS signal will now be investigated mathematically. A brief review of the treatment of frequency modulation by a single sine wave tone will add to the clarity of the discussion; this treatment can be found in any fundamental text on radio engineering, a reference is given for convenience.[5] The usual

(5) Goldman, op. cit., pp. 141-179.

presentation begins with defining the instantaneous angular frequency in a frequency modulated system by

$$(24)\quad \frac{1}{2\pi}\frac{d\theta}{dt} = f_c + (\Delta f)\sin \omega_0 t$$

In this expression, $f_c$ is the frequency of the carrier being modulated; ( $\Delta f$ ) is the frequency deviation and depends upon modulator design; $f_o$ is the modulating frequency. The expression (24) is then integrated, and the function of time resulting from frequency modulation is taken as

$$(25)\quad f(t) = A\sin\theta = A\sin\left[2\pi f_c t - \frac{\Delta f}{f_o}\cos\omega_0 t + \varphi\right]$$

The final expression (25) can then be expanded by Bessel's function identities to obtain the signal spectrum.

An extension of the above procedure is needed in order to analyse frequency modulation by a square wave. For multitone modulation, the instantaneous frequency can be defined as

$$(26)\quad \frac{1}{2\pi}\frac{d\theta}{dt} = f_c + \sum (\Delta_K f)\sin K\omega_0 t$$

In this expression, $f_c$ is the frequency of the carrier; $f_o$ is the fundamental frequency component of the keying wave; ( $\Delta_K f$) is the deviation caused by the kth component of the keying wave. In frequency modulation, the deviation caused by a modulating component is directly proportional to the amplitude of that component. For a square wave modulating signal, the ( $\Delta_K f$) can be replaced by ( $\frac{\Delta f}{K}$ ) where ( $\Delta f$) is arbitrary and will depend upon modulator design.

$$(27) \quad \frac{1}{2\pi} \frac{d\theta}{dt} = f_c + \sum_{K=1,3,5,\cdots}^{\infty} \frac{\Delta f}{K} \sin K\omega_0 t$$

When (27) is integrated, the result including the constant of integration is

$$(28) \quad \theta = \omega_c t - \sum \frac{\Delta f}{K^2 f_0} \cos K\omega_0 t + C_1$$

$$(29) \quad f(t) = A \sin\left[(\omega_c t + C_1) - \sum \frac{\Delta f}{K^2 f_0} \cos K\omega_0 t\right]$$

In (29) the constant $C_1$ which has the nature of a phase angle has been lumped with the carrier component. It will be later shown that $C_1$ can be set to zero. Now the summation is the Fourier expansion of the saw tooth wave shown in Figure 7. The saw tooth wave will have the same period as the modulating signal, since it was obtained by integrating the keying wave. The various lines making up the saw tooth wave form can be expressed as $F(t) = \pm at \pm b_n$. This means that the summation in (29) is replaceable by the equation of a line in appropriate intervals and that (29) itself will break up into functions of the type

$$f(t) = A \sin(\omega_c t + C_1 \pm at \pm b_n)$$

, each defined in an appropriate interval. However, the slopes and intercepts of the lines must be investigated, keeping in mind the condition of continuity of the sine functions at the transition points. The expansion of a saw tooth wave in (29) has as its leading coefficient $\frac{\Delta f}{f_0}$ where $\Delta f$ is arbitrary and would depend upon a particular modulator design. So the amplitude of the wave has been taken arbitrarily as E. From the assumed amplitude, the known period, and known points on the lines, the slopes and intercepts of each can be calculated quite simply. The results are tabulated below.

$$
\begin{array}{lll}
(30) & F(t) = \frac{2E}{T}t - E & T > t > 0 \\
 & F(t) = -\frac{2E}{T}t + 3E & 2T > t > T \\
 & F(t) = \frac{2E}{T}t - 5E & 3T > t > 2T \\
 & F(t) = -\frac{2E}{T}t + 7E & 4T > t > 3T \\
 & \quad \cdots\cdots\cdots\cdots & 
\end{array}
$$

When these results are substituted into (29), the result is: ($C_1$ has been set to zero for simplicity; this step will be justified later.)

$$
\begin{array}{lll}
(31) & f(t) = A \sin\left(W_c t + \frac{2Et}{T} - E\right) & T > t > 0 \\
 & f(t) = A \sin\left(W_c t - \frac{2E}{T}t + 3E\right) & 2T > t > T \\
 & f(t) = A \sin\left(W_c t + \frac{2Et}{T} - 5E\right) & 3T > t > 2T \\
 & f(t) = A \sin\left(W_c t - \frac{2E}{T}t + 7E\right) & 4T > t > 3T \\
 & \quad \cdots\cdots\cdots\cdots & 
\end{array}
$$

Now that the equations of the pulses of radio frequency in an FS signal are known, we can see if they satisfy the continuity condition as required. The second and third of the above equations should match at t = 2T. To test, we ask if

$$(32) \qquad W_c\, 2T - \frac{2E}{T}(2T) + 3E = W_c(2T) + \frac{2E}{T}(2T) - 5E$$

$$E = E$$

Further tests will show that the functions are matched at each transition point as required. The matching is independent of $C_1$ since it affects each pulse equally; furthermore, the matching condition reduces to an identity in E so that E, which is related to an arbitrary frequency shift, is itself entirely arbitrary. Since $\frac{2E}{T}$ has the dimensions of frequency, we can set $\omega_c + \frac{2E}{T} = \omega_1$ and $\omega_c - \frac{2E}{T} = \omega_2$ ; E has the nature of a phase constant, permitting the spectrum functions to be written in the following compact form:

$$(33) \quad \begin{aligned} f(t) &= A \sin\left[\omega_1 t - (4k+1)\phi\right] & (2k+1)T > t > 2kT \\ f(t) &= A \sin\left[\omega_2 t + (4k+3)\phi\right] & (2k+2)T > t > (2k+1)T \end{aligned}$$

In the above set of functions, the first two pulses to the right of the origin are accounted for by k = 0; the next two by k = 1, and so on. It must be stressed that the phase angle $\phi$ is not arbitrary, except insofar as it is related to the arbitrary frequency shift, and cannot be set to zero. Furthermore, it has been retained in the functions (33) because they satisfy the necessary condition that the pulses be joined at transition points. The matching came about quite naturally from frequency modulation theory and <u>was in no way forced.</u> It is possible, of course, for certain combinations of frequency deviation and keying frequency to cause $\phi$ to be an even number of $\pi$ radians, in which case it is effectively zero in (33). This special case will be considered later.

The signal should be examined for periodicity. (We are still concerned only with periodicity based on a single keying interval.) The presence of the phase angles is enough to make one suspect that periodicity does not exist. Before testing for periodicity, it is necessary to evaluate the constant E in terms of the signal parameters. The leading term of the Fourier expansion of a triangular wave of amplitude E is $\frac{4E}{\pi}$ , the leading term of the expansion dealt with here is $\Delta f/f_o$. Equating and solving, there results $E = \phi = \frac{\pi \Delta f}{4 f_o}$ . It is sufficient for the present purposes to test for a particular case rather than to formulate a general condition in terms of one of the k's in (33). Take for example $f(\frac{T}{2})$ which should be equal to $f(\frac{5T}{2})$. Substituting in (33), there results as the requirement for periodicity

$$(34) \quad \sin\left(\omega_1 \frac{T}{2} - \frac{\pi \Delta f}{4 f_o}\right) = \sin\left(\omega_1 \frac{5T}{2} - \frac{5\pi \Delta f}{4 f_o} \pm 2n\pi\right)$$

which is reducible to

$$(35) \quad n = \frac{f_1}{f_o} - \frac{\Delta f}{2 f_o}$$

The condition required by (35) is not satisfied except for special choices of $\omega_1$ , $\Delta$ f, and $f_o$, so that we may conclude that periodicity of the signal is a special case rather than the general one. To summarize briefly at this point, it has been shown that the result of frequency modulating a carrier frequency $f_c$ with a square wave of fundamental $f_o$ is to produce a signal which consists of alternate pulses of frequencies $f_1$ and $f_2$; the pulses are joined at the transition points from $f_1$ to $f_2$; periodicity of the signal does not exist except as a special case; the frequencies $f_1$ and $f_2$ are respectively above and below $f_c$ by the same amount. It can be seen that four cases might arise: (1) the pulses of $f_1$ and $f_2$ start with their natural phases, and periodicity does not exist; (2) periodicity exists, but the pulses do not start with their natural phases; (3) the pulses start with their natural phases, and periodicity also exists; (4) the pulses do not start with their natural phases, and periodicity does not exist. Each case will give rise to a different type of spectrum which will be discussed in turn.

Case 1. The pulses start with their natural phases, and periodicity does not exist.

The mathematical requirement in this case is that the phase angle $\phi$ be an even multiple of $\pi$ radians since a shift of $2n\pi$ radians is physically equivalent to no shift at all. Note that this condition requires that the pulse beginning at the origin start with 0. Physically, this could be taken care of by adjusting the timing of the keying or translating the axes to such a pulse. Translation of axes will not affect the nature of the signal and its spectrum. Mathematically, then, it is required that

(36) $\phi = \dfrac{\pi \Delta f}{4 f_0} = 2n\pi$ or $\dfrac{\Delta f}{8 f_0} = n$

Since periodicity does not occur in this case, Fourier analysis based on a keying interval would not be correct. However, the signal can be regarded as two interlocking waves, each amplitude modulated with a square wave. The pulses of $f_1$ and $f_2$ start with their natural phases so that application of amplitude modulation theory is perfectly rigorous. The spectrum is then obtained by superimposing two functions of the type (19). The $f_2$ spectrum should include a phase factor of T to account for the relative displacement of the $f_1$ and $f_2$ pulses, but since the amplitudes and frequencies involved are the major point of interest in the signal spectrum, it is sufficient to omit phase angles and write

$$(37) \quad S(t) = \frac{A}{2}\sin\omega_1 t + \sum\frac{A}{\pi k}\cos(\omega_1 - k\omega_0)t - \sum\frac{A}{\pi k}\cos(\omega_1 + k\omega_0)t$$
$$+ \frac{A}{2}\sin\omega_2 t + \sum\frac{A}{\pi k}\cos(\omega_2 - k\omega_0)t - \sum\frac{A}{\pi k}\cos(\omega_1 + k\omega_0)t$$

Case 2. Periodicity exists, but the pulses do not start with their natural phases.

Non-periodicity in the general case has been previously demonstrated in order to show the nature of the signals dealt with. However, at this time, the conditions for periodicity should be examined closely, bearing in mind that it must be satisfied for successive $f_2$ pulses as well as $f_1$ pulses. Remembering that a particular k in (33) accounts for a pair of pulses, an analysis similar to that used to develop (36) and applied to kth and (k + 1)th pairs of pulses will yield, the k dropping out

(38) $n = \frac{f_1}{f_0} - \frac{\Delta f}{2 f_0}$ and $p = \frac{f_2}{f_0} + \frac{\Delta f}{2 f_0}$

In the above expressions n and p are integers which come from adding angles of $2n\pi$ & $2p\pi$ as in the derivation of (36). Adding the above expressions and changing n+p, which is an integer to the integer n, the condition for periodicity is obtained.

(39) $n = \frac{f_1 + f_2}{f_0}$

The spectrum for this case has been obtained by Balth Van der Pol. However, the various special cases treated in this article are not treated in the work of Van der Pol. He obtained the spectrum of the signal without regard to the constants which amount to phase shift, and his work would be applicable in any case to a signal in which pure frequency modulation without phase shift is considered to be the end result. Such a signal would be difficult to obtain physically inasmuch as the constants provide for continuity of the signal, and continuity would seem to be required in any circuit containing inductance. The spectrum obtained by Van der Pol in the notation of this article follows:

(40) $$f(t) = \frac{2A}{\pi}\Big[\frac{m}{m^2}\sin\frac{\pi m}{2}\cos\omega t$$
$$+\frac{m}{m^2-1^2}\cos\frac{\pi m}{2}\left\{\cos(\omega-\omega_0)t-\cos(\omega+\omega_0)t\right\}$$
$$-\frac{m}{m^2-2^2}\sin\frac{\pi m}{2}\left\{\cos(\omega-2\omega_0)t+\cos(\omega+2\omega_0)t\right\}$$
$$-\frac{m}{m^2-3^2}\cos\frac{\pi m}{2}\left\{\cos(\omega-3\omega_0)t-\cos(\omega+3\omega_0)t\right\}$$
$$+\frac{m}{m^2-4^2}\sin\frac{\pi m}{2}\left\{\cos(\omega-4\omega_0)t+\cos(\omega+4\omega_0)t\right\}$$
$$-\cdots\cdots\Big]$$
$$m = \frac{\Delta f}{f_0} \qquad f_c = \frac{f_1+f_2}{2}$$

<u>Case 3.</u> The signal is periodic, and the pulses start with their natural phases.

The mathematical conditions for this case have already been derived; it is required that conditions (36) and (39) be satisfied simultaneously, so that

(41) $\frac{f_1+f_2}{f_0} = n \quad , \quad \frac{\Delta f}{8f_0} = p$ where n and p are integers.

Since both conditions are satisfied in this case, we would expect (37) and (40) to give the same result although they are of somewhat different form. Since the frequency shift in this case is a harmonic of $f_0$, and the condition $\frac{f_1+f_2}{f_0} = n$ is also satisfied, it follows that $f_c$, $f_1$, $f_2$ are each also harmonics of $f_0$. Thus, each sideband of (37) differs from either $f_c$, $f_1$, $f_2$ by an integer harmonic of $f_0$; and for each frequency of the part of the spectrum which is centered on $f_1$, there is an equal frequency in the $f_2$ spectrum, enabling one to rearrange the counting to start with $f_c$. Consider, by way of illustration, the nth sideband above $f_c$, and presume also that it is the kth sideband above $f_1$. Evidently for the $f_1$ spectrum, k is given by n minus $\Delta f/f_0$, and the same frequency will be found in the $f_2$ spectrum when k equals $\Delta f/f_0$ plus n. The total term for this frequency component is then

$$(42)\quad -\frac{A}{\pi\left(n-\frac{\Delta f}{f_0}\right)}\cos\left[\omega_1+\left(n-\frac{\Delta f}{f_0}\right)\omega_0\right]t-\frac{A}{\pi\left(\frac{\Delta f}{f_0}+n\right)}\cos\left[\omega_2+\left(\frac{\Delta f}{f_0}+n\right)\omega_0\right]t$$

which becomes

$$(43)\quad -\frac{2A}{\pi}\,\frac{\frac{\Delta f}{f_0}}{\left(\frac{\Delta f}{f_0}\right)^2-n^2}\cos\left(\omega_c+n\omega_0\right)t$$

This last expression is the nth upper sideband of the Van der Pol spectrum since $\cos\frac{\Delta f}{f_0}\frac{\pi}{2}=1$ because $\frac{\Delta f}{f_0}$ is an even integer and $\sin\frac{\Delta f}{f_0}\frac{\pi}{2}$ is zero for the same reason. For this special case, then, the spectrum of the signal is given equally well by either (37) or (40).

Case 4. The signal pulses do not start with their natural phases, and the signal is not periodic.

This last case is, of course, the one most likely to occur in practice. Neither of the spectra thus far derived is rigorously applicable, and there is no elementary method of determining mathematically what the spectrum really is. Fourier integral analysis can usually be employed to find the spectrum of a non-periodic function provided the integral of the function from minus to plus infinity is finite and determinate. A signal of infinite extent and oscillatory in nature such as the one being analysed here cannot be attacked directly, if at all, by the Fourier integral.

The physical appearance of the signal in this case is very little different from that of Case 1. It would appear reasonable to consider (37) as an approximation to the signal spectrum. Since the signal consists of a high frequency oscillation, it will be almost periodic, and it would seem just as reasonable to accept the Van der Pol spectrum as an approximation. However, since neither is rigorously applicable, one would have no right to set either up as a standard over the other.

### 3. Comparison of the Derived Spectra for an FS Signal

A signal corresponding to Case 4 of Section 2 has been chosen for plotting and comparison of (37) and (40); the plotted spectra are shown in Figures 8 and 9. The values of the parameters chosen are typical of those that might be encountered in radio teletype practice, a carrier frequency of 6 mc/s, a keying frequency of 23 c/s, and a frequency shift of 425 c/s. At best, the plots are approximations to the spectrum of the particular signal under

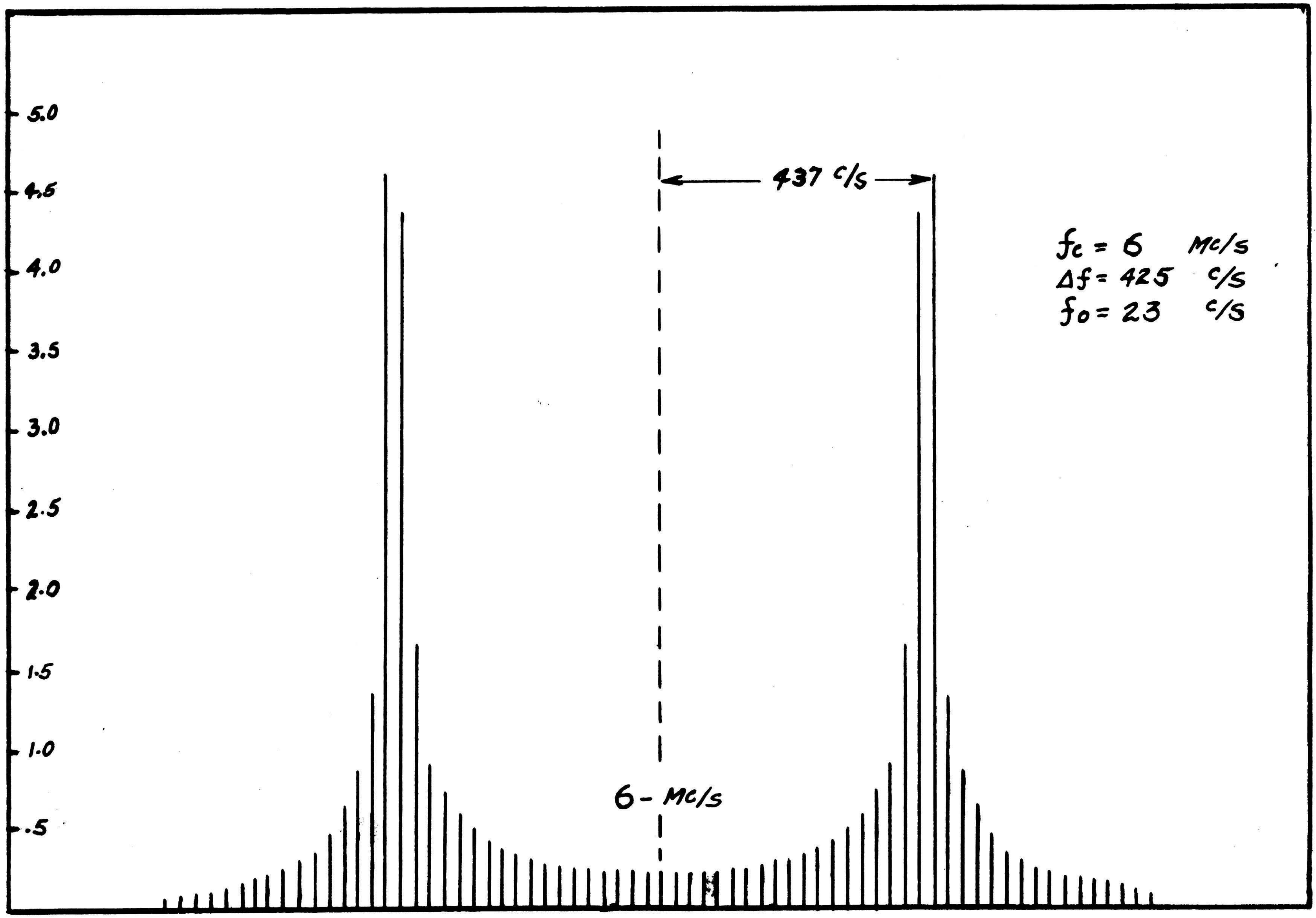


FIG 8. SPECTRUM OF FS KEYED SIGNAL, VAN DER POL THEORY

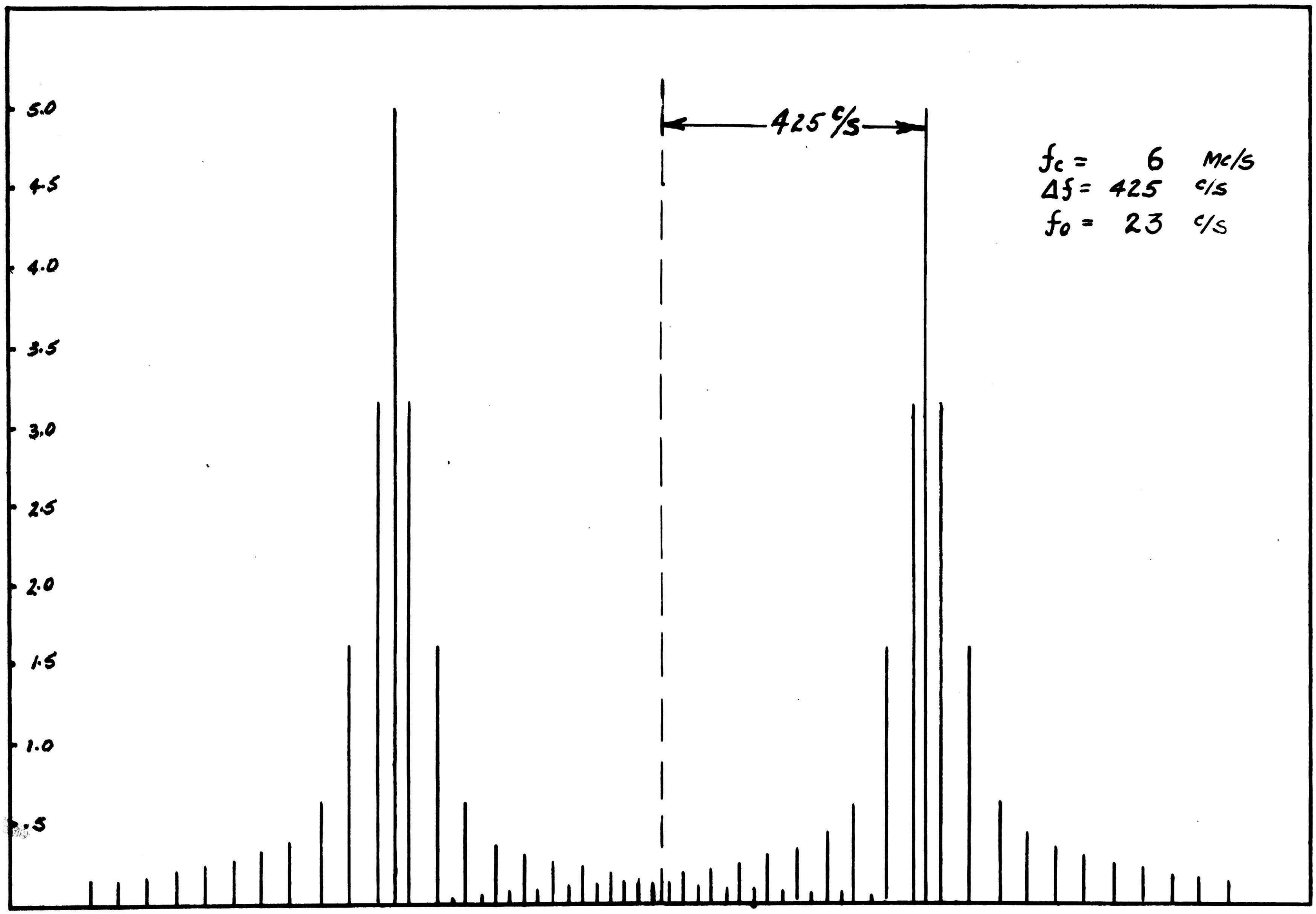


FIG 9. SPECTRUM OF FS KEYED SIGNAL - AMPLITUDE MODULATION THEORY

study. The line frequencies in Figure 8 are not the same as those in Figure 9 and should not be taken very seriously in either case. The general outline, maximum amplitudes, distance between peaks, and bandwidth are properties of the spectra which are of more importance.

The bandwidth of (37) is easily computed. Following a common practice in radio engineering, the band will be taken as including all components which are not more than 40 DB below the unmodulated carrier. The unmodulated carrier is of amplitude A since when the transmitter is not being keyed, the output is a wave of amplitude A and frequency $f_1$; in practice, a transmitter rests on steady "mark" during idle periods. We can solve for the component which is 40 DB below the carrier as follows:

$$(44) \quad 40 = 20 \, Log \frac{A}{A/\pi k}$$

The component satisfying this equation to the nearest whole number is the 31st, but it must be remembered that the counting of k is with respect to components which are displaced 425 cps from the center of the spectrum and that there are two such kth components, one below $f_1$, and the other above $f_2$. The bandwidth is then

$$(45) \quad B_1 = 2\left[23(31) + 425\right] = 2276 \ c/s$$

Computation of the bandwidth for the Van der Pol spectrum is somewhat more tedious. The decibel relation between the amplitude of the unmodulated carrier and the nth sideband can be written as

$$(46) \quad Z = \log \left| \frac{A}{\frac{2A\,m\theta}{m^2-n^2}} \right| \qquad \theta = \cos\frac{m\pi}{2} \text{ or } \sin\frac{m\pi}{2}$$

The largest value of n which is found to satisfy the above equation for the chosen parameters is 34; hence, the bandwidth is given by

$$(47) \quad B_2 = 2\,[34(23)] = 1564 \text{ c/s}$$

With regard to the very important detail of bandwidth, the two spectra of (37) and (40) give widely different results. It is common practice in FS keyed transmitters to employ a keying filter to remove some of the higher order harmonics of the keying wave, thereby reducing the actual transmitted bandwidth.[(6)] Whether or not the bandwidth

(6) Sprague, Robert M., Frequency Shift Radio Teletype and Telegraph System, Electronics, pp. 127-131, November 1944.

requirement is actually reduced by such a keying filter is a question which cannot be answered by the techniques employed in this paper, since the spectrum without the filter is not rigorously given by either (37) or (40). It is to be presumed that manufacturers of FS keying equipment have experimental verification of the benefit in bandwidth reduction to be obtained by using a keying filter. However, there seems to be no simple analytical method of computing the benefit.

# CONCLUSIONS

The conclusions to this paper are to be found in the derived signal spectra (19), (37), and (40), each applicable as noted under various possible cases in the Discussion. No rigorous spectrum distribution was found for the most general case of FS keying, and it was found that when (37) and (40) are considered separately as approximations to the spectrum in the general case, widely different results were obtained. The question as to which is the better approximation cannot be answered unless the actual spectrum distribution is known precisely, either through experiment or a more extended analysis. Analysis should probably proceed on the basis of the Fourier integral, although it is not directly applicable to the problem in a simple manner. A second possibility for extension of the theory is through joint use of amplitude, frequency, and phase modulation theory since it was shown that frequency modulation of a carrier by a square wave results in phase shift by an amount necessary to insure continuity of the resulting signal. Any extension of the results of this paper should also include the special case where periodicity of the signal can be based on more than one keying interval.

# BIBLIOGRAPHY

1. Books:

Goldman, Stanford, Frequency Analysis, Modulation and Noise, McGraw-Hill Book Co., pp. 141-179, 1948.

Terman, F. E., Radio Engineering, McGraw-Hill Book Co., p. 493, 1947.

2. Periodicals

Buff, C., Frequency Shift Keying Techniques, Radio, Press Wireless Mfg. Corp., vol. 30, no. 8, August 1946.

Peterson, Atwood, others, Observations and Comparison on Radio Telegraph Signalling by FS and On-Off Keying, RCA Review, vol. 7, no. 1, March 1946.

Rea, W. T., Effect of Telegraph Distortion on Margin of Operation of Start-Stop Receiver, Bell System Tech. Journal, vol. 62, no. 12, July 1944.

Sprague, Robert M. Frequency Shift Radio Teletype and Telegraph System, Electronics, pp. 127-131, November 1944.

Watson, E. F., Fundamentals of Teletype Used in the Bell System, Bell System Tech. Journal, p. 620, October 1938.

3. Publications of Learned Societies

Van der Pol, Balth, Frequency Modulation, Proc. I. R. E., vol. 18, pp. 1194-1205, July 1930.

## VITA

The author was born in St. Louis, Missouri, on August 13, 1922. He attended grade school at Bonne Terre, Missouri, and high school at Flat River, Missouri, being graduated from the latter in 1941. After graduation from high school, he attended the University of Cincinnati until October of 1942 when he withdrew to enter the US Army Air Force. He served as an enlisted radio technician and then as an instructor in the Air Force Technical School at Sioux Falls, South Dakota. He was commissioned as a communications officer in April of 1944 and assigned to duty with the Army Airways Communications in North Africa until being separated from the service in June of 1946. After being separated from the service, he worked for the MacKay Radio and Telegraph Company at Tangier, Morocco, as a radio teletype transmitter and receiver technician. The author returned to the United States and entered the Missouri School of Mines and Metallurgy in September 1947. He was graduated with a B.S. in E.E. in July 1949, and since that time has been doing graduate work in Physics while holding a graduate assistantship in E.E. during the academic year 1949-50 and an instructorship in Engineering Mechanics during 1950-51.

The author was married June 1946 while in Tangier, Morocco, to Helene Amzallag. Katherine Nora Duncan became a member of the family in January 1949.