Letting $\bar{t} = \mu t$, $x = \mu j$ this simplifies to

$$1 - (B^*W_2)(t) \sim \frac{(1 - A_2)}{2\pi A_1 a_2^2} \sum_{x=1}^{t-2} \frac{e^{-a_1 x - a_2(\bar{t} - x - 1)}}{x^{3/2}(\bar{t} - x - 1)^{3/2}}$$

where $-a_i = \ln A_i + 1 - A_i$, $i = 1,2$.

Note that conditions 1 and 2 are negligible in comparison with 3. However, one could safely form the sum of the three conditions.

References

[1] J. Riordan, *Stochastic Service Systems*. New York: Wiley, 1962.
[2] D. Fraser, *Statistics—An Introduction*. New York: Wiley, 1958.
[3] R. G. Miller, Jr., "Priority Queues," *Ann. Math. Stat.*, vol. 31, pp. 86–103, 1960.

Application of the Maximum Principle to the Design of Minimum Bandwidth Pulses

D. M. DiTORO, STUDENT MEMBER, IEEE, AND K. STEIGLITZ, MEMBER, IEEE

***Abstract*—The minimum bandwidth pulse for a given rms value and peak amplitude is found using the Pontryagin Maximum Principle for bounded phase space. The solution is obtained by actually solving the equations of the maximum principle rather than by a verification of a solution arrived at by other means. The resulting optimal pulse shape can be used for comparisons with more easily generated pulses, and may be considered to be the optimal modification of a rectangular pulse from the point of view of minimum bandwidth. Pulses of this sort are useful in pulse communication systems.**

I. Introduction

The PROBLEM of finding the best pulse shape to use in a pulse communication system has been the object of investigation for some time. At the turn of the century A. C. Crehore and G. O. Squier [1] suggested that the rectangular pulses then being used on the Atlantic Cable be replaced by half-cycle sine pulses. They claimed that the half-cycle sine pulse was more efficient than the rectangular pulse because, for a given peak value of voltage, more energy is transmitted through the pass band of the cable.

In 1924, H. Nyquist [2] argued that the half-cycle sine pulse was inferior to a rectangular pulse which had been suitably filtered. From the point of view of received energy and spectral spread, his proposed pulse was superior. His reasoning, however, was based on the principle that the shape of the received signal was practically independent of the shape of the pulses used in the transmitted signal. In modern pulse communication systems, this principle does not apply.

Manuscript received March 29, 1965; revised June 21, 1965. This work was partially supported by the National Science Foundation and by the U.S. Army Research Office under Contract DA-31-124-ARO-D-292. Computer facilities used were partially sponsored by the National Science Foundation Grant NSF-GP 579.

The authors are with the Department of Electrical Engineering, Princeton University, Princeton, N.J.

The half-cycle sine pulse appeared again in 1946 when D. Gabor [3] showed that of all pulses of finite duration and given energy the half-cycle sine pulse had the minimum bandwidth. He arrived at the solution using the classical calculus of variations. His result was as follows: for a fixed rms value defined by

$$E^2_{rms} = \frac{1}{T} \int_0^T p^2(t)dt \tag{1}$$

and a bandwidth defined by

$$B^2 = T \int_0^T (dp/dt)^2 dt, \tag{2}$$

the pulse which minimizes this bandwidth is

$$p(t) = \sqrt{2}E_{rms} \sin(\pi t/T), \quad 0 \leq t \leq T. \tag{3}$$

J. H. H. Chalk [4] in 1950 derived a set of optimal pulse shapes using an alternate criterion. He considered pulse shapes which maximized the ratio of available energy to total energy of the pulse, where available energy is that remaining after a linear filtering of the pulse. The pulses he obtained using this criterion have discontinuities at the endpoints and therefore have spectra which are similar asymptotically to the rectangular pulse. In fact, according to (2), they have infinite bandwidth.

A simpler approach was taken by Z. Jelonek and E. Fitch [5], who proposed a modification of the rectangular pulse. First a trapezoidal pulse was considered, and then the upper corners of this pulse were rounded. They showed that the rise time of the pulses at the receiver was not unduly diminished, while the bandwidth was decreased markedly.

This intuitive approach to the design of a suitable pulse is interesting because it reflects qualities of an optimal pulse which heretofore have not been incorporated in the mathematical formulation. In fact the modification of rectangular pulses to reduce their bandwidth and yet maintain their desirable qualities, such as low peak-to-rms ratio and small rise time, suggests the introduction of an amplitude constraint into Gabor's formulation. The following problem, then, will be considered in this paper: What pulse shape of fixed finite duration and energy, with amplitude constrained to be less than or equal to a fixed quantity, has the minimum bandwidth as given by (2)?

This pulse shape must approach a rectangular pulse as the peak-to-rms ratio approaches unity, since this is the only possible shape with unity peak-to-rms ratio. Also it must reduce to the half-cycle sine pulse when the peak-to-rms ratio is $\sqrt{2}$, since the half-cycle sine pulse satisfies the stated requirements and has a peak-to-rms ratio of $\sqrt{2}$. For intermediate values of peak-to-rms ratio, the solution represents, from the point of view of minimum bandwidth, the optimum modification of a rectangular pulse. The solution to this problem provides a guide to the trade-off between bandwidth and peak-to-rms ratio of pulses and leads to an inequality involving these parameters.

From a mathematical point of view this problem is of interest because the classical calculus of variations no longer applies. The problem can be formulated as an optimal control problem in bounded phase space, and the Pontryagin Maximum Principle [6] can be used to obtain necessary conditions for optimality. It turns out that these conditions uniquely specify a trajectory in the phase space and hence, uniquely determine the pulse shape.

In a recent paper, D. W. Tufts and D. A. Shnidman [7] considered the situation in which a peak-value constraint is added to the matched filter problem of radar theory. In their application of the maximum principle, the unknown pulse was the control variable and the peak-value constraint bounded the control region. The solution was found to be a clipped version of the previously known solution. In the formulation that follows, the unknown pulse appears as a phase coordinate and the amplitude constraint bounds the phase space. It is interesting to note that a clipped version of the solution given by (3) is not the optimum pulse shape for the problem considered in this paper.

II. Formulation

The Pontryagin Maximum Principle can be applied to a system described by a vector differential equation of the form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}),$$

where $\mathbf{x}$ is the n-vector of phase coordinates and $\mathbf{u}$ is the r-vector of control variables [8]. The dot will be used to denote a time derivative. In order to cast the proposed problem in this form let the phase coordinate $x_1(t) = p(t)$, the unknown pulse shape. Let $x_2(t)$ correspond to the accumulated energy in the pulse at time t, and let $x_3(t)$ be a time coordinate. Finally, let $u(t)$, the control, be the derivative of $x_1(t)$. Time is included as a phase coordinate and the time t_1 when the trajectory intersects the terminal point $\mathbf{x}(t_1)$ will be thought of as unknown. This puts the problem in the form considered by Pontryagin [9].

The following differential equations incorporate these definitions:

$$\begin{aligned} \dot{x}_1(t) &= u(t) \\ \dot{x}_2(t) &= x_1^2(t) \\ \dot{x}_3(t) &= 1. \end{aligned} \tag{4}$$

The initial conditions at the time $t = 0$ are obtained as follows: $x_1(0) = 0$ since any discontinuity gives an infinite bandwidth; $x_2(0) = 0$ since the energy of the pulse at the start of the trajectory is zero; and $x_3(0) = 0$ since the pulse starts at $t = 0$. The final conditions at t_1 are: $x_1(t_1) = 0$ for the same reason as the initial condition on x_1; $x_2(t_1) = TE^2_{rms}$ the required total energy; and $x_3(t_1) = T$ the time interval given.

The appropriate control problem is: Find the control $u(t)$ such that the functional

$$\int_0^{t_1} u^2 dt$$

is minimized. This corresponds to minimum bandwidth since $u(t) = \dot{x}_1(t)$. Following Pontryagin [10] an additional phase coordinate $x_0(t)$ is defined so that

$$\dot{x}_0(t) = u^2(t) \tag{5}$$

with the initial condition $x_0(0) = 0$. The coordinate $x_0(t)$ is proportional to the bandwidth of the pulse when $t = T$. Finally the constraint on the maximum value of the pulse becomes

$$|x_1(t)| \le E_m, \quad 0 \le t \le t_1.$$

This constraint defines an allowable region in the phase space and to such a situation a more complicated form of the maximum principle applies [11]. Let G be the closed region in the phase space defined by the constraint $|x_1| \le E_m$. Then the segments of the optimal trajectory interior to G satisfy the maximum principle as given in Section III. The segments on the boundary of G satisfy a more complicated form of the maximum principle given in Section IV. At each junction point, the jump condition described in Section V is satisfied. It will be seen that the required pulse shape follows as a direct consequence of these conditions.

III. The Maximum Principle in the Interior of G

For the system of differential equations given by (4) and (5) the Hamiltonian $H(\mathbf{x}, \boldsymbol{\psi}, u)$ which is defined [12] to be $H(\mathbf{x}, \boldsymbol{\psi}, u) = \dot{\mathbf{x}} \cdot \boldsymbol{\psi}$ becomes

$$H(\mathbf{x}, \boldsymbol{\psi}, u) = \psi_0 u^2 + \psi_1 u + \psi_2 x_1^2 + \psi_3, \tag{6}$$

where $\boldsymbol{\psi}$ is the state vector of the adjoint system. The differential equations for the adjoint system are obtained from the Hamiltonian equation [12]

$$\frac{d\psi_i}{dt} = -\frac{\partial H}{\partial x_i}, \; i = 0, \ldots, 3.$$

That is

$$\begin{aligned}\dot{\psi}_0 &= 0\\ \dot{\psi}_1 &= -2x_1\psi_2\\ \dot{\psi}_2 &= 0\\ \dot{\psi}_3 &= 0.\end{aligned}\tag{7}$$

For constant values of ψ and $\mathbf{x}$, the Hamiltonian $H(\mathbf{x}, \psi, u)$ is explicitly a function of the control u. Define [13] $M(\psi, \mathbf{x})$ as the least upper bound of $H(\mathbf{x}, \psi, u)$ over all allowable controls u. The control which achieves this least upper bound is the optimal control. In the general situation to which the maximum principle can be applied, restrictions on the control can be introduced. In this situation, however, no restriction on $u = \dot{x}_1$ has been made. Therefore, the ordinary theory of maxima and minima applies. Setting

$$\frac{\partial H}{\partial u} = 0$$

requires that

$$u = -\psi_1/2\psi_0 \tag{8}$$

and therefore,

$$M(\psi, \mathbf{x}) = -\frac{\psi_1^2}{4\psi_0} + \psi_2 x_1^2 + \psi_3. \tag{9}$$

The condition that the stationary point is indeed a maximum, namely that $\partial^2 H/\partial u^2 < 0$ implies that $\psi_0 < 0$.

Solutions can now be given for $\mathbf{x}$ and ψ using the expression for the optimal control u, (8), obtained above. These solutions will be denoted by $\mathbf{x}^-$ and ψ^-, indicating they are interior to G. There must be such a segment of optimal trajectory because the initial condition on $\mathbf{x}$ is within G. The solutions to the differential equations are:

$$\begin{aligned}x_0^-(t) &= x_0^-(0) + \int_0^t (\psi_1^-/2\psi_0^-)^2 dt\\ x_2^-(t) &= x_2^-(0) + \int_0^t (x_1^-)^2 dt\\ x_3^-(t) &= x_3^-(0) + t\\ \psi_0^-(t) &= \psi_0^-\\ \psi_2^-(t) &= \psi_2^-\\ \psi_3^-(t) &= \psi_3^-.\end{aligned}\tag{10a}$$

The simultaneous differential equations (4) and (7) for x_1 and ψ_1

$$\begin{aligned}\dot{x}_1 &= -\psi_1/2\psi_0\\ \dot{\psi}_1 &= -2x_1\psi_2\end{aligned}$$

have the solutions

$$\begin{aligned}x_1^-(t) &= A\sin(\omega t) + B\cos(\omega t)\\ \psi_1^-(t) &= 2\psi_0^-\omega B\sin(\omega t) - 2\psi_0^-\omega A\cos(\omega t)\end{aligned}\tag{10b}$$

where

$$\omega^2 = -\psi_2^-/\psi_0^-. \tag{11}$$

Here A and B are arbitrary constants.

If it is assumed that the final conditions on $\mathbf{x}$ can be satisfied without a segment of optimal trajectory on the boundary of G, then the arbitrary constants in the foregoing solutions can be evaluated immediately. For such a final condition, however, the restriction on the phase space is superfluous. Since there is no restriction on the control, the problem reduces to one in the classical calculus of variations [14], and the solution is the half-cycle sine pulse as previously indicated.

In the general situation some of the arbitrary constants in the above solution are determined from the maximum principle in the interior of G and the initial condition on $\mathbf{x}$. The remaining constants are evaluated in conjunction with the segment of optimal trajectory on the boundary of G. The initial condition $\mathbf{x}(0)$ requires that

$$\begin{aligned}x_0^-(0) &= 0\\ B &= 0\\ x_2^-(0) &= 0\\ x_3^-(0) &= 0.\end{aligned}\tag{12}$$

The maximum principle [13], in addition to requiring that u be such that $H(\mathbf{x}, \psi, u)$ is maximum in the variable u, further stipulates that

$$\psi_0^- \le 0$$

and that

$$M[\psi^-(t), \mathbf{x}^-(t)] = 0.$$

This later condition requires that

$$A^2\psi_2^- + \psi_3^- = 0 \tag{13}$$

which is obtained by substituting the solutions for $\mathbf{x}$ and ψ [(10)] into the expression for $M(\psi, \mathbf{x})$ [(9)].

All the conditions of the maximum principle in the interior of G have been accounted for, and the next step is to consider the segment of the optimal trajectory on the boundary of G.

IV. The Maximum Principle on the Boundary of G

The form of the maximum principle [11] that will be applied is more complex than the previous case, and some preliminary definitions will be necessary [15]. The allowable region G in the phase space is to be defined by a set of the form

$$G = \{\mathbf{x}: g(\mathbf{x}) \le 0\}$$

in the neighborhood of the boundary. Since this definition is meant to apply only near the boundary of G, and only one boundary need be considered at a time as x_1 is either positive or negative, it will be assumed that x_1 reaches the positive boundary $x_1 = E_m$. The solution is identical if the negative boundary is considered, since the energy and bandwidth involve the square of x_1. Hence, for positive values of x_1, let

$$g(\mathbf{x}) = x_1 - E_m. \tag{14}$$

The function $g(\mathbf{x})$ has continuous second partial derivatives as required and

$$\text{grad } g(\mathbf{x}) = (0, 1, 0, 0) \tag{15}$$

is nonzero on the boundary of G.

The function $p(\mathbf{x}, u)$ is defined [16] to be the dot product of the normal vector to the boundary of G [(15)] and the velocity of the phase point $\mathbf{x}$:

$$p(\mathbf{x}, u) = \text{grad } g(\mathbf{x}) \cdot \dot{\mathbf{x}}.$$

In order for the phase point to remain on the boundary of G, $p(\mathbf{x}, u)$ must be zero; in this case the phase point has no velocity component normal to the boundary of G. From (4) and (5), which specify the velocity of the phase point, and from (15) for the gradient of $g(\mathbf{x})$, $p(\mathbf{x}, u)$ is found to be

$$p(\mathbf{x}, u) = u. \tag{16}$$

A phase point $\mathbf{x}$ is called regular with respect to a control point u_1 if

$$p(\mathbf{x}, u_1) = 0$$

and

$$\frac{\partial p(\mathbf{x}, u_1)}{\partial u} \neq 0.$$

Let $\omega(\mathbf{x})$ be the set of all u for which $\mathbf{x}$ is regular. That is:

$$\omega(\mathbf{x}) = \{u : \mathbf{x} \text{ is regular}\}.$$

The trajectory $\mathbf{x}(t)$ corresponding to $u(t)$ and lying on the boundary of G is also called regular if $u(t) \in \omega x(t)$. Only these trajectories will remain on the boundary of G; hence only these trajectories need be considered.

Define $m(\boldsymbol{\psi}, \mathbf{x})$ for $\mathbf{x}$ on the boundary of G by

$$m(\boldsymbol{\psi}, \mathbf{x}) = \sup_{u \in \omega(\mathbf{x})} H(\mathbf{x}, \boldsymbol{\psi}, u). \tag{17}$$

For $p(\mathbf{x}, u)$ given by $p = u$ [(16)] the set $\omega(\mathbf{x})$ has the one member $u = 0$. Therefore $m(\boldsymbol{\psi}, \mathbf{x})$ is found by substituting the optimal control $u = 0$ into $H(\boldsymbol{\psi}, \mathbf{x}, u)$ [(6)].

For trajectories on the boundary of G the differential equations for the adjoint system involve an additional term which does not appear in the previous case. This term is the product of a Lagrange multiplier $\lambda(t)$ defined by

$$\frac{\partial H(\mathbf{x}, \boldsymbol{\psi}, u)}{\partial u} = \lambda(t) \frac{\partial p(\mathbf{x}, u)}{\partial u} \tag{18}$$

and the factor

$$\frac{\partial p(\mathbf{x}, u)}{\partial \mathbf{x}}. \tag{19}$$

In the present situation, $p(\mathbf{x}, u)$ depends only on u [(16)], and the term given by (19) vanishes. Hence the differential equations for $\mathbf{x}$ and $\boldsymbol{\psi}$ are those given previously [(4), (5), (6)] with the optimal control $u = 0$. The solutions will be denoted by $\mathbf{x}^+$ and $\boldsymbol{\psi}^+$, indicating they lie on the boundary of G. These are

$$\begin{aligned} \overset{+}{x}_0(t) &= \overset{+}{x}_0 \\ \overset{+}{x}_1(t) &= \overset{+}{x}_1 \\ \overset{+}{x}_2(t) &= \overset{+}{x}_2(\tau) + \int_\tau^t (\overset{+}{x}_1) dt \\ \overset{+}{x}_3(t) &= \overset{+}{x}_3(\tau) + t, \end{aligned} \tag{20}$$

and

$$\begin{aligned} \overset{+}{\psi}_0(t) &= \overset{+}{\psi}_0 \\ \overset{+}{\psi}_1(t) &= \overset{+}{\psi}_1(\tau) - 2\overset{+}{x}_1 \overset{+}{\psi}_2 (t - \tau) \\ \overset{+}{\psi}_2(t) &= \overset{+}{\psi}_2 \\ \overset{+}{\psi}_3(t) &= \overset{+}{\psi}_3, \end{aligned} \tag{20}$$

where τ is the start of the (+) segment of the trajectory.

The remaining necessary conditions that follow from the maximum principle are

$$m[\boldsymbol{\psi}^+(t), \mathbf{x}^+(t)] = 0 \tag{21}$$

$$\overset{+}{\psi}_0 \leq 0 \tag{22}$$

$$\boldsymbol{\psi}^+(\tau) \neq 0 \tag{23}$$

$$\boldsymbol{\psi}^+(\tau) \cdot \text{grad } g(\mathbf{x}(\tau)) = 0 \tag{24}$$

$$\frac{d\lambda}{dt} \leq 0. \tag{25}$$

As indicated previously $m(\boldsymbol{\psi}, \mathbf{x})$ [(17)] is found by substituting $u = 0$ into the Hamiltonian. Hence evaluating (21) yields

$$\overset{+}{\psi}_2 (\overset{+}{x}_1)^2 + \overset{+}{\psi}_3 = 0. \tag{26}$$

The next two conditions [(22) and (23)] give no relationships between the constants, although they limit their possible values and can be checked from the final solution. The fourth condition [(24)] requires that $\overset{+}{\psi}_1(\tau) = 0$. The Lagrange multiplier $\lambda(t)$ is found from its defining equation (18). Remembering that $u = 0$, the fifth condition [(25)] requires that

$$\frac{d\lambda}{dt} = \frac{d\overset{+}{\psi}_1}{dt} = -2\overset{+}{x}_1 \overset{+}{\psi}_2 \leq 0. \tag{27}$$

That is

$$\overset{+}{\psi}_2 \leq 0$$

as $\overset{+}{x}_1$ is assumed positive.

The initial condition of $\mathbf{x}^+$ is determined by the endpoint of $\mathbf{x}^-$, since the trajectory must be continuous. The remaining initial conditions for $\boldsymbol{\psi}^+$ are obtained from the jump condition.

V. The Jump Condition and the Final Solution

Let $\mathbf{x}(\tau)$ be the junction point of $\mathbf{x}^-(t)$ and $\mathbf{x}^+(t)$ on the boundary of G. The jump condition [17] is satisfied if the corresponding adjoint vectors are related by the following equation

$$\psi^{+}(\tau) = \psi^{-}(\tau) + \mu \operatorname{grad} g[\mathbf{x}(\tau)] \tag{28}$$

for arbitrary μ. In order to facilitate the evaluation of the remaining parameters, the solutions [(10) and (11)] will be rewritten incorporating the relationships among the constants so far encountered:

$$\begin{aligned} \mathbf{x}^{-}(t) &= \left(\int_0^t (-\psi_1^{-}/2\psi_0^{-})^2 dt,\ A \sin(\omega t),\ \int_0^t (x_1^{-})^2 dt,\ t\right) \\ \psi^{-}(t) &= (\psi_0^{-},\ -2\psi_0^{-}\omega A \cos(\omega t),\ \psi_2^{-},\ -A^2\psi_2^{-}) \end{aligned} \tag{29}$$

and

$$\begin{aligned} \mathbf{x}^{+}(t) &= \left(x_0^{+},\ x_1^{+},\ x_2^{+}(\tau) + \int_\tau^t (x_1^{+})^2 dt,\ x_3^{+}(\tau) + t\right) \\ \psi^{+}(t) &= (\psi_0^{+},\ -2x_1^{+}\psi_2^{+}(t-\tau),\ \psi_2^{+},\ -(x_1^{+})^2\psi_2^{+}). \end{aligned} \tag{29}$$

For $\mathbf{x}^{-}$ and ψ^{-}, what remains to be determined is A and $\omega^2 = -\psi_2^{-}/2\psi_0^{-}$. Since the Hamiltonian H is homogeneous in the components of ψ, these components need only be determined up to a multiplicative constant [18]. This is reflected in the fact that ω is the significant parameter. For $\mathbf{x}^{+}$ the initial condition is obtained from $\mathbf{x}^{-}$ at $t = \tau$ the junction time. That is

$$\mathbf{x}^{-}(\tau) = \mathbf{x}^{+}(\tau). \tag{30}$$

In particular:

$$x_1^{+} = E_m \tag{31}$$

as this is the boundary of G. For ψ^{+} the jump condition (28) gives the following relationships

$$\begin{aligned} \psi_0^{-} &= \psi_0^{+} \\ \psi_2^{-} &= \psi_2^{+} \\ -A^2\psi_2^{-} &= -(x_1^{+})^2\psi_2^{+}. \end{aligned}$$

This must be the case since grad $g(\mathbf{x})$ [(15)] has a nonzero component corresponding to only ψ_1; hence all other components of ψ must be continuous at the junction. In particular these equations imply that

$$A = E_m. \tag{32}$$

Therefore what remains to be found is ω, which will specify the junction time τ. Since ψ_2^{+} must be positive [(27)] and ψ_0^{+} must be negative [(22)], ω is real.

The final conditions on $\mathbf{x}(t)$, namely

$$\begin{aligned} x_1(t_1) &= 0 \\ x_2(t_1) &= TE^2_{rms} \\ x_3(t_1) &= 0, \end{aligned}$$

require that x_1 leave the boundary of G and return to $x_1 = 0$. It is clear that this segment is entirely analogous to the segment $\mathbf{x}^{-}$ in that the same equations and solutions apply. The jump condition at the departure point requires that the amplitude and frequency of the sinusoid be the same as those of the $(-)$ trajectory. The final condition for $x_3(t_1)$ gives t_1, the duration of the trajectory. As before, the frequency ω specifies the departure time. Hence what remains is the endpoint of $x_2(t_1)$, that is, the energy constraint. This specifies the frequency ω, the only remaining unknown parameter.

The solution for $x_1(t)$ is therefore

$$\begin{aligned} x_1(t) &= E_m \sin(\omega t) && 0 \le t \le \pi/2\omega \\ &= E_m && \pi/2\omega \le t \le T - \pi/2\omega \\ &= E_m \sin \omega(T-t) && T - \pi/2\omega \le t \le T \end{aligned} \tag{33}$$

VI. An Inequality

The rms value and bandwidth of the optimum pulse (33) are found to be

$$\left.\begin{aligned} E^2_{rms} &= E_m^2\left(\frac{1+\alpha}{2}\right) \\ B^2 &= E_m^2\left(\frac{\pi^2}{2(1-\alpha)}\right), \end{aligned}\right\} \quad 0 \le \alpha < 1$$

where α is the fraction of the interval $[0,T]$ during which the pulse is at its maximum value. If the bandwidth and peak value are normalized by the rms energy and α is eliminated from the preceeding equations, the following equality is obtained for the optimum pulse shape:

$$\frac{B}{E_{rms}} = \frac{\pi}{2}\,\frac{\left(\dfrac{E_m}{E_{rms}}\right)^2}{\sqrt{\left(\dfrac{E_m}{E_{rms}}\right)^2 - 1}}, \qquad 1 \le \frac{E_m}{E_{rms}} \le \sqrt{2}$$

Since all other pulses of finite duration have a larger bandwidth for a given peak-to-rms ratio, the following inequality results for all pulses of finite duration:

$$\frac{B}{E_{rms}} \ge \frac{\pi}{2}\,\frac{\left(\dfrac{E_m}{E_{rms}}\right)^2}{\sqrt{\left(\dfrac{E_m}{E_{rms}}\right)^2 - 1}}, \qquad 1 \le \frac{E_m}{E_{rms}} \le \sqrt{2},$$

where the definitions (1) and (2) have been used. For

$$\frac{E_m}{E_{rms}} \ge \sqrt{2}$$

the optimal trajectory does not intersect the boundary and the solution is given by the half-cycle sine pulse. In this case the inequality becomes simply

$$\frac{B}{E_{rms}} \ge \pi, \qquad \frac{E_m}{E_{rms}} \ge \sqrt{2}.$$

These inequalities correspond to well-known uncertainty relationships when the amplitude constraint is removed and when different definitions of bandwidth or duration are used. For a discussion of these results the reader is referred to the work of Landau and Pollak [19].

REFERENCES

[1] A. C. Crehore and G. O. Squier, "A practical transmitter using the sine wave for cable telegraphy; and measurements with alternating currents upon an atlantic cable," *Trans. AIEE*, vol. XVII, pp. 385–431, May 18, 1900.
[2] H. Nyquist, "Certain factors affecting telegraph speed," *Trans. AIEE*, vol. XLIII, pp. 412–422, February 1924.
[3] D. Gabor, "Theory of communication," *J. IEE (London)*, vol. 93, pt. III, pp. 429–457, November 1946.
[4] J. H. H. Chalk, "The optimum pulse-shape for pulse communication," *Proc. IEE (London)*, vol. 97, pt. III, pp. 88–92, March 1950.
[5] D. Cooke, Z. Jelonek, A. J. Oxford, and E. Fitch, "Pulse communication," *J. IEE (London)*, vol. 94, pt. IIIA, pp. 83–105, March 1947.
[6] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Michchanko, *The Mathematical Theory of Optimal Processes*. New York: Interscience, 1962, p. 1.
[7] D. W. Tufts and D. A. Shnidman, "Optimum waveforms subject to both energy and peak-value constraints," *Proc. IEEE*, vol. 52, pp. 1002–1007, September 1964.
[8] L. S. Pontryagin et al., *The Mathematical Theory of Optimal Processes*. New York: Interscience, sec. 2, p. 11.
[9] *Ibid.*, sec. 8, p. 66.
[10] *Ibid.*, sec. 2, p. 13.
[11] *Ibid.*, ch. VI, p. 257.
[12] *Ibid.*, sec. 3, p. 18.
[13] *Ibid.*, sec. 3, p. 19.
[14] *Ibid.*, ch. V, p. 239.
[15] *Ibid.*, sec. 31, p. 258.
[16] *Ibid.*, sec. 32, p. 264.
[17] *Ibid.*, sec. 36, p. 302.
[18] *Ibid.*, sec. 4, p. 22.
[19] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—II," *Bell Sys. Tech. J.*, vol. 40, pp. 65–84, January 1961.

Frequency Division in Speech Bandwidth Reduction

R. E. BOGNER

***Abstract*—The following phenomena are discussed in relation to experimental observations on a number of frequency division schemes: "division" of a wave containing several frequencies; phase ambiguity; and strobing. Phase ambiguity is believed to be a fundamental limitation, and the results of a number of previous experiments are attributed to its effects.**

I. Introduction

During the past decade a number of attempts have been made to reduce the bandwidth required for the transmission of speech by the use of "frequency division" techniques. In most cases, the experiments have produced distortions which were not adequately explained. It is the purpose of this article to discuss some basic ideas about "frequency division" which enable the phenomena observed in the various experiments to be explained.

II. Previous Observations

It is of interest to examine the schemes which have been tried in the past, together with the hypotheses on which they were based, and the results observed. This review is necessarily brief and omits many of the results which do not have a bearing on frequency division phenomena.

a) The experiments of Marcou and Daguet [1] were initiated by the fact that "constant amplitude" speech is highly intelligible [2], [3]. A single sideband (carrier suppressed) signal is expressible in the form

$$f(t) = A(t) \cos \Phi(t)$$

where $A(t)$ can be thought of as the instantaneous amplitude or envelope, and $\cos \Phi(t)$ as the oscillation filling in the envelope.

Figure 1 (a)–(c) illustrates the process. This representation is valid and useful for baseband signals (i.e., carrier frequency of zero), but the physical picture is not so clear then. A "constant amplitude" signal may then be represented by $\cos \Phi(t)$. This signal may be generated by amplitude clipping $f(t)$ [Fig. 1(d)] and then filtering to remove harmonic components. The "instantaneous frequency," $\Phi'(t) = d\Phi(t)/dt$ [Fig. 1(e)], may be obtained by frequency measurement.

The bandwidth compression scheme, shown in Fig. 2, was based on the idea of reducing the excursions of $\Phi'(t)$ by "frequency division" [Fig. 1(e)]. The "frequency divided" signal, $\cos \Phi(t)/n$, was demonstrated to have a narrower effective spectrum than $\cos \Phi(t)$, as was expected from frequency modulation theory. When the frequency was "multiplied" again with no filtering, the result was highly intelligible. Then a filter was interposed between the frequency divider and the frequency multiplier, as would be necessary to remove minor extraneous components for narrow-band transmission. The resultant frequency multiplied signal was then seriously distorted.

b) The schemes of Bogert [4] and the first scheme of Bogner and Smith [5] were similar. It had been noticed that the three main vowel formants in speech usually occurred in the frequency ranges 200–1000, 1000–2000 and 2000–3200 c/s, respectively. Accordingly, as in

Manuscript received December 10, 1964.
The author is with the Department of Electrical Engineering, University of Queensland, St. Lucia, Brisbane, Queensland, Australia.