

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XLIV

APRIL 1965

NUMBER 4

Copyright 1965, American Telephone and Telegraph Company

Automatic Equalization for Digital Communication

By R. W. LUCKY

(Manuscript received November 24, 1964)

Distortion in transmission channels causes intersymbol interference in digital communication systems. This distortion may be partially corrected at the receiver through the use of a tapped delay line having adjustable tap gain settings (transversal filter). The problem of minimizing distortion with a finite-length transversal filter is examined. In the region of small initial channel distortion where most existing systems operate, the best tap gain settings satisfy a set of simultaneous linear equations. For larger initial distortion, iterative techniques are required to find best gain settings. The distortion is shown to be a convex function of the tap gains, so mathematical programming techniques may be employed for optimization.

The practical problem is that of evolving a logical strategy whereby the tap gains of the transversal filter may be set to optimum values. An easily implemented system for automatic equalization is described which makes use of a steepest-descent technique of minimization. The equalizer is automatically set prior to data transmission in a training period during which a series of test pulses is transmitted. Only polarity information is required, so digital logic may be used in the equalizer. For application to high-speed data transmission, great accuracies are required for the tap gain settings. Thus the problem of noise in the channel during equalization is quite important. The final error due to noise and channel distortion and the equalizer settling time are discussed and evaluated. Finally, the effect of a transversal filter equalizer in terms of the system frequency-domain characteristics is considered.

I. INTRODUCTION

Present data rates on voice telephone channels are limited to about 2400 bits per second. Although the noise margin on these facilities is sufficient to permit much higher rates, the nonuniform transmission characteristics of the channel cause what might be termed a distortion barrier, prohibiting faster transmission. The distortion of data pulses by the channel results in these pulses being smeared out in time so as to overlap other transmitted pulses. This intersymbol interference is one of the chief degrading factors in present systems and becomes the determining factor in the design of higher-rate systems. To alleviate the effects of intersymbol interference it is necessary to equalize the channel.

In the past equalization has generally been accomplished by flattening the amplitude characteristic and linearizing the phase characteristic using fixed amplitude-frequency and phase-frequency networks. Although this type of equalization is adequate for speech transmission requirements, it does not provide the precise control over the channel time response which is necessary for high-speed data transmission. Thus to realize the full transmission capability of the channel there is a need for automatic, time-domain equalization.

Among the basic philosophies for automatic equalization of data systems are pre-equalization at the transmitter and post-channel equalization at the receiver. Since the former technique requires a feedback channel, we will concentrate our efforts here on equalization at the receiver. This equalization can be performed either during a training period prior to data transmission or it can be performed continuously during data transmission. The typical voice channel changes little during an average data call, so pre-call equalization should be sufficient in most cases. Many of the principles and techniques which will be discussed here can be applied to the continuous or adaptive equalizer, although our main concern will be in the pre-call automatic equalizer.

Most engineers are agreed that multilevel vestigial sideband transmission offers the best hope for higher-speed transmission on voice channels. There are good theoretical reasons for this choice of modem on a channel of limited bandwidth and high signal-to-noise ratio. The equalization problem we will discuss is based on the use of VSB transmission or what is equivalent, baseband transmission. The equalizer is to be placed at the receiver directly after the demodulation process. Thus as far as the equalizer is concerned the transmission is baseband.

Suppose now that the equivalent baseband system transmits amplitude a_n at time nT , where a_n is chosen from a set of M possible discrete amplitudes. The single-pulse response of the over-all system, including

channel and equalizer, we call $h(t)$. The received signal $y(t)$ is then

$$y(t) = \sum_{n=-\infty}^{\infty} a_n h(t - nT). \quad (1)$$

At some suitably chosen sampling time t_0 the output voltage $y(t_0)$ may be abbreviated

$$y_0 = \sum_{n=-\infty}^{\infty} a_n h_{-n} \quad (2)$$

where

$$h_n = h(t_0 + nT). \quad (3)$$

This voltage is the sum of the wanted term a_0 plus an intersymbol interference term

$$y_0 = h_0 \left[a_0 + \frac{1}{h_0} \sum_{n=-\infty}^{\infty}{}' a_n h_{-n} \right]. \quad (4)$$

(The prime will be used very frequently on summations to indicate deletion of the $n = 0$ term.)

Now in the second term of (4) the a_n coefficients are chosen by the data system user according to some probabilistic rule. Since in a VSB system the maximum positive and negative of values of a_n are equal in magnitude, say a_{\max} , the maximum value the intersymbol interference term can assume is

$$\text{max interference} = \frac{a_{\max}}{h_0} \sum_{n=-\infty}^{\infty}{}' |h_n|. \quad (5)$$

We thus define an interference criterion proportional to (5) called distortion and labelled D

$$D = \frac{1}{h_0} \sum_{n=-\infty}^{\infty}{}' |h_n|. \quad (6)$$

The so-called eye opening for an M -level VSB system is simply and monotonically related to D .

$$I = 1 - (M - 1)D. \quad (7)$$

The eye opening is a widely accepted criterion of a data system's performance. In what follows we shall always use the equalizer to minimize the distortion D defined by (6). A particular advantage of this criterion is that it is not dependent on noise statistics or on the statistical distribution of the customer's data sequence $\{a_n\}$. It is a minimax

criterion in that we seek to maximize the customer's minimum margin against noise over all data sequences. A heuristic argument can be made for a criterion using the sum of the squares of the h_n 's. Obviously neither criterion minimizes average probability of error—a mathematically intractable problem even when assuming Gaussian noise and independent, equally likely input symbols. Practically speaking the minimization of either criterion leads to negligibly different results.

Now what we need is a variable filter for an equalizer which can exercise wide control over the time response samples $\{h_n\}$. The transversal filter shown in Fig. 1 is ideal for this purpose.^{1,2} This filter consists of a continuous delay line tapped at T -second intervals. Each tap has an associated attenuator which in combination with an inverter is capable of giving variable gain. The filter output is the sum of all the attenuated tap voltages. Such a filter is capable of flexible control of the output time sequence $\{h_n\}$ when used in tandem with the channel to be equalized. In fact, under suitable conditions, derived in the final section of this paper, a transversal filter of infinite length may be used to completely eliminate distortion in a channel. Generally speaking, a transversal filter of finite length cannot eliminate distortion, so we will

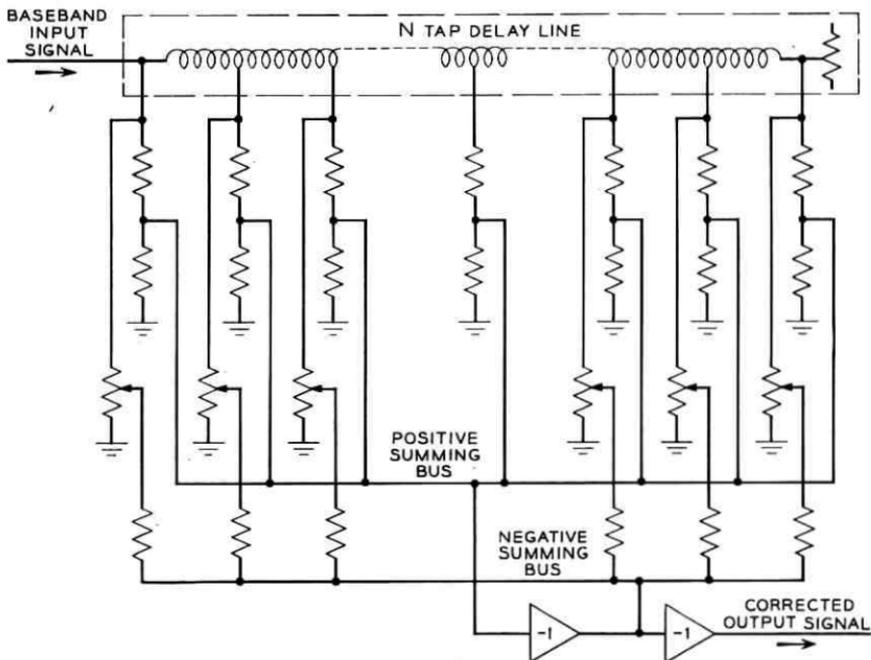


Fig. 1—Elements of a transversal filter.

study the minimization of distortion D with a transversal filter of finite length. We then show an implementation of an automatic equalizer which achieves minimum distortion under certain conditions generally satisfied on voice telephone channels.

In succeeding sections the theoretical and practical limitations of the equalizer as implemented are studied. In order to effectively eliminate distortion in voice channels great accuracies are required in the setting of the tap gain coefficients. Background noise and settling time for the equalizer become extremely important. Finally we discuss the behavior of the transversal equalizer, essentially a time-domain device using a time-domain criterion, in the frequency domain.

II. THE MINIMIZATION OF DISTORTION

We will assume that the transversal filter has $N + 1$ taps with associated tap gains. One of these taps is taken as the reference with its gain denoted c_0 , while the other N taps are placed somewhere at integer multiples of T seconds before or after tap c_0 . The positions of these taps are usually the locations $-(N/2)T, -(N/2 + 1)T, \dots, (N/2)T$, but since these particular locations are not necessary to any of the theorems we use a more general formulation. Let K_N be the set of integers denoting the positions of the $N + 1$ tap locations

$$K_N = \{n \mid \text{tap location exists at time } nT \text{ from reference}\}.$$

The tap gains will be denoted $c_j, j \in K_N$.

The impulse response at the input to the transversal filter is denoted $x(t)$ and its samples at times nT form the time sequence $\{x_n\}$. We will assume for convenience that this response is normalized so that $x_0 = 1$. Thus the distortion in the pulse $x(t)$ prior to equalization, called initial distortion D_0 , is

$$D_0 = \sum_{n=-\infty}^{\infty} |x_n|. \quad (8)$$

The transversal filter serves as a device to multiply time sequences. In this case the input sequence $\{x_n\}$ is multiplied by the tap gain sequence $\{c_n\}$ according to the rules of polynomial multiplication. It can be seen by inspection that the output sequence $\{h_n\}$ is formed by the rule

$$h_n = \sum_{j \in K_N} c_j x_{n-j}. \quad (9)$$

The final distortion we seek to minimize is

$$D = \frac{1}{h_0} \sum_{n=-\infty}^{\infty} |h_n|. \quad (10)$$

The reference response sample h_0 causes somewhat of a practical problem. In a multilevel system the slicing levels are generally fixed so that the gain must be closely controlled. In other words it is necessary to provide a normalizing control to adjust h_0 to unity, the assumed value of x_0 . This can be done two ways which lead to slightly different mathematical problems.

Problem 1:

Fix the tap c_0 at unity. Minimize the criterion D over the N variables $\{c_j\}; j \in K_N, j \neq 0$. The output pulse may be normalized if desired by an over-all gain control outside the transversal filter.

Problem 2:

Let the tap c_0 be variable. Minimize the criterion D over the $N + 1$ variables $\{c_j\}, j \in K_N$, subject to the constraint $h_0 = 1$.

A study of these two problems reveals that the minimum distortion in either case is the same. Except in isolated cases ($c_0 = 0$ in Problem 2) the optimum tap gains in Problems 1 and 2 are related by a constant factor. There is a practical difference in the range of tap gains required and a mathematical difference in the type of function involved. (Problem 1 is nonlinear, while 2 is piecewise linear.) Since the minima are the same and the optimum gains related, we concentrate on Problem 2 because of its simpler properties.

The constraint on h_0 may be written

$$h_0 = 1 = \sum_{j \in K_N} c_j x_{-j}. \quad (11)$$

Solving for the gain c_0 we have

$$c_0 = 1 - \sum_{j \in K_N} c_j x_{-j}. \quad (12)$$

We substitute (12) into (9) to obtain

$$h_n = \sum_{j \in K_N} c_j (x_{n-j} - x_n x_{-j}) + x_n. \quad (13)$$

Now since h_0 is unity the distortion (10) becomes

$$D = \sum_{n=-\infty}^{\infty} \left| \sum_{j \in K_N} c_j (x_{n-j} - x_n x_{-j}) + x_n \right|. \quad (14)$$

Now that c_0 has been eliminated to satisfy the constraint on h_0 we desire to minimize D in (14) over the N variables c_j ; $j \in K_N, j \neq 0$.

First observe that D is a continuous, piecewise-linear function of the variables $\{c_j\}$. We can rewrite (14) in the form

$$D = \sum'_{j \in K_N} c_j \sum'_{n=-\infty}^{\infty} (x_{n-j} - x_n x_{-j}) \operatorname{sgn} h_n + \sum'_{n=-\infty}^{\infty} x_n \operatorname{sgn} h_n \quad (15)$$

$$\operatorname{sgn} h_n = \begin{cases} +1, & h_n \geq 0 \\ -1, & h_n < 0. \end{cases}$$

In this equation the coefficients of the c_j 's are constant over certain regions of the N -dimensional space of definition of $\{c_j\}$. Breakpoints where the coefficients assume new values occur whenever an output sample h_n is zero. A minimum cannot occur between breakpoints where the function is linear; thus at least one value h_{k_1} is zero at the minimum. The equation $h_{k_1} = 0$ may be used to eliminate one of the N variables c_i . The reduced equation is of the same piecewise linear form, requiring at least one more output sample $h_{k_2} = 0$, etc. We arrive at the conclusion that at least N samples of the output time sequence $\{h_n\}$ must be zero at the minimum. But N equations of the form $h_{k_i} = 0$; $i = 1, \dots, N$ are sufficient to determine the values of the tap gains $\{c_j\}$. We need only solve N simultaneous linear equations using (13) with $n = k_i$; $i = 1, \dots, N$.

The question remains as to the values of the k_i 's, i.e., which N zeros in the output response sequence does one force to achieve minimum distortion? In most cases of interest this question is answered by the following theorem, which is proved in Appendix A.

Theorem I

If $D_0 < 1$, then the minimum distortion D must occur for those N tap gains which simultaneously cause $h_n = 0$ for all $n \in K_N, n \neq 0$.

Another important property of the distortion function which is both useful and descriptive is the following:

Theorem II

If the tap c_0 is used to satisfy the constraint $h_0 = 1$, then the distortion D is a convex function of the N variables c_j ; $j \in K_N; j \neq 0$.

Proof

For convenience denote settings of the equalizer by the N -component vectors $\bar{\alpha}$ and $\bar{\sigma}$. To prove convexity of D it is necessary to show that

for any two settings $\bar{\alpha}$ and $\bar{\sigma}$ and for all λ , $0 \leq \lambda \leq 1$,

$$D[\lambda\bar{\alpha} + (1 - \lambda)\bar{\sigma}] \leq \lambda D(\bar{\alpha}) + (1 - \lambda)D(\bar{\sigma}). \quad (16)$$

This equation would show that the distortion always lies on or beneath a chord joining values of distortion in N -space.

From (14)

$$D[\lambda\bar{\alpha} + (1 - \lambda)\bar{\sigma}] = \sum_{n=-\infty}^{\infty} \left| \sum_{j \in K_N} [\lambda\alpha_j + (1 - \lambda)\sigma_j] \cdot (x_{n-j} - x_n x_{-j}) + x_n \right| \quad (17)$$

$$D[\lambda\bar{\alpha} + (1 - \lambda)\bar{\sigma}] = \sum_{n=-\infty}^{\infty} \left| \lambda \left\{ \sum_{j \in K_N} \alpha_j (x_{n-j} - x_n x_{-j}) + x_n \right\} + (1 - \lambda) \left\{ \sum_{j \in K_N} \sigma_j (x_{n-j} - x_n x_{-j}) + x_n \right\} \right| \quad (18)$$

$$D[\lambda\bar{\alpha} + (1 - \lambda)\bar{\sigma}] \leq \lambda D(\bar{\alpha}) + (1 - \lambda)D(\bar{\sigma}). \quad (19)$$

One of the most important properties of convex functions is that they possess no relative minima other than their absolute minimum. Thus any minimum of D found by systematic search or other mathematical programming methods must be the absolute (or global) minimum of distortion.

In summary, we have shown that D is a continuous, piecewise-linear, convex function of the N tap gains. This function has a single minimum which must occur when N zeros appear in the output time sequence $\{h_n\}$. If the initial distortion is less than 100 per cent ($D_0 < 1$), then the minimum occurs when the N samples of the output time sequence which correspond in location to the N taps on the transversal filter are simultaneously zero. This description is illustrated in two simple cases in Figs. 2 and 3. In Fig. 2 only one tap is variable ($N = 1$), while in Fig. 3 equal distortion contours are plotted for an example of a 2-tap equalizer. In both cases the initial distortion is less than 100 per cent, so the minima are easily located.

III. EQUALIZATION STRATEGIES AND IMPLEMENTATION

3.1 Strategy When $D_0 < 1$

The condition $D_0 < 1$, which is sufficient for easy location of the minimum distortion, is equivalent to the condition that the unequalized channel is capable of supporting binary transmission without error in the absence of noise. It may be seen from (7) that $D_0 = 1$ implies a

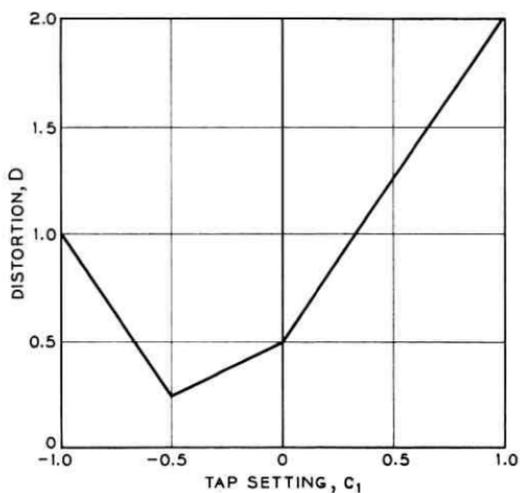


Fig. 2—Distortion vs tap setting in a one-dimensional example.

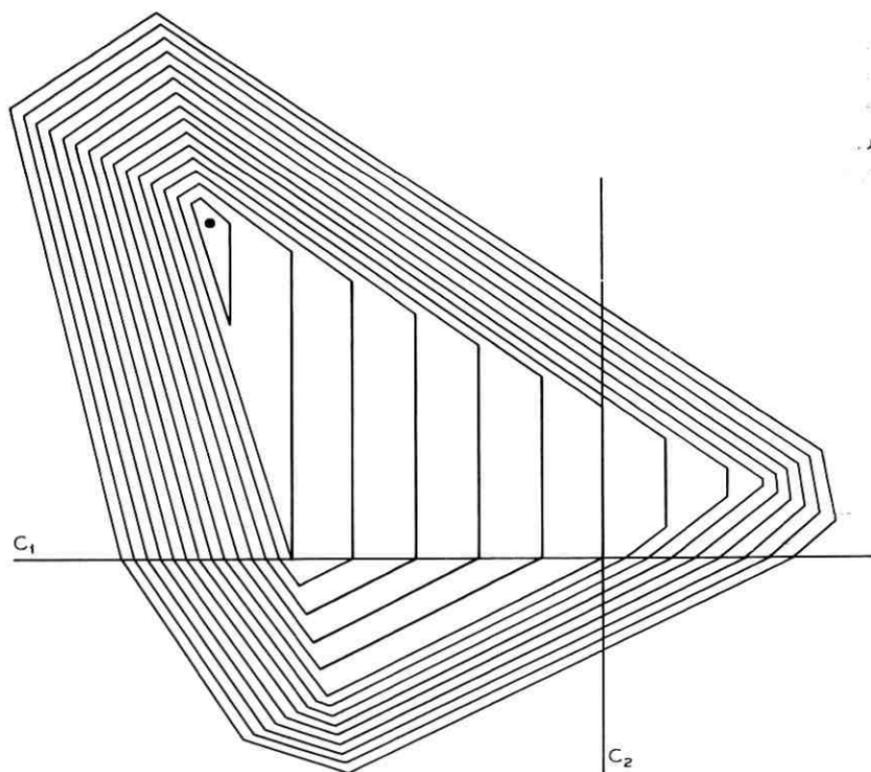


Fig. 3—Equal distortion contours for a two-tap example.

completely closed binary eye. In almost all cases of interest on voice telephone channels the modem will support binary transmission without equalization, and so the condition is met. In these cases equalization enables multilevel operation to take advantage of a relatively high signal-to-noise ratio. For the present we will deal with the design of an equalizer where $D_0 < 1$.

To make matters more concrete we will henceforth assume that the N adjustable taps of the transversal filter are divided equally before and after the reference tap. In the several equalizer models which have been built the normalization has been carried out sometimes by an outside gain control and sometimes by an adjustable center tap. In either case the task of the automatic tap gain setting apparatus is to zero the N output samples h_n ; $|n| \leq N/2$, $n \neq 0$. By Theorem I this achieves minimum distortion. One can, of course, derive the tap gains by the simultaneous solution of N linear equations, but from an instrumentation point of view there are many simpler schemes. The condition $D_0 < 1$ ensures a "loosely coupled" system where the interaction between tap gains is weak. Thus there are a number of iterative strategies which converge to the desired settings. In all these schemes a sequence of test pulses is transmitted prior to actual data transmission. After each test pulse the tap gains of the equalizer are readjusted in such a way as to eventually result in the proper N zeros in the output time sequence. The choice between such schemes is dictated by ease of instrumentation, settling time required for equalization and accuracy of the final settings.

It is true that with a special-purpose computer the optimum tap settings could be computed using only a single test pulse; however, this overlooks the presence of background and impulse noise on the facility. Each test pulse is in itself unreliable, so many pulses must be averaged in some way to give accurate settings. The equalization system must also be relatively unaffected by any large bursts or impulses which occur during the setup period.

An equalization strategy has been devised which meets all requirements and is easily instrumented. The motivation for this strategy is based on a steepest-descent technique. Consider the use of an outside gain control for normalization with the center tap gain fixed at unity. Solving the N simultaneous equations $h_n = 0$ for $|n| \leq N/2$, $n \neq 0$ is equivalent to minimizing the "truncated distortion"

$$D_N = \sum_{n=-N/2}^{N/2} |h_n|. \quad (20)$$

The function D_N is a convex function amenable to solution by steepest-descent techniques. After each test pulse has been received the tap gains are incremented so that the N -dimensional incrementing vector is in a direction opposed to the gradient of D_N . This gradient may be written

$$\bar{\nabla} D_N = \sum_{j=-N/2}^{N/2} \frac{\partial D_N}{\partial c_j} \bar{a}_j \quad (21)$$

where \bar{a}_j is a unit vector in the direction of the c_j coordinate. The components of the gradient are

$$\frac{\partial D_N}{\partial c_j} = \sum_{n=-N/2}^{N/2} \frac{\partial h_n}{\partial c_j} \text{sgn } h_n \quad (22)$$

$$\frac{\partial D_N}{\partial c_j} = \sum_{n=-N/2}^{N/2} x_{n-j} \text{sgn } h_n. \quad (23)$$

Now we approximate (23) by assuming the samples x_n , for $n \neq 0$, are small in comparison to x_0 , which is unity.

$$\frac{\partial D_N}{\partial c_j} \approx \text{sgn } h_j \quad (24)$$

$$\bar{\nabla} D_N \approx \sum_{j=-N/2}^{N/2} \text{sgn } h_j \bar{a}_j. \quad (25)$$

The approximation has resulted in an extremely simple expression, since all steps are of equal magnitude and the direction of each step is determined by simply taking the polarity of the corresponding output sample. No analog voltages are involved, so digital logic can be used in the tap gain setting circuitry. Before describing this circuitry, it is necessary to demonstrate that this iterative scheme does indeed converge to the desired minimum. After each test pulse, each tap gain c_j is incremented by an amount $-\Delta \text{sgn } h_j$, so that the new output samples are

$$h_n^* = \sum_{j=-N/2}^{N/2} (c_j - \Delta \text{sgn } h_j) x_{n-j} + x_n \quad (26)$$

$$h_n^* = h_n - \Delta \sum_{j=-N/2}^{N/2} \text{sgn } h_j x_{n-j} \quad (27)$$

$$h_n^* = h_n - \Delta \text{sgn } h_n - \Delta \sum_{\substack{j=-N/2 \\ j \neq n}}^{N/2} \text{sgn } h_j x_{n-j} \quad (28)$$

$$|h_n^*| \leq ||h_n| - \Delta| + \Delta D_0. \quad (29)$$

Since $D_0 < 1$, we have

$$|h_n^*| < ||h_n| - \Delta| + \Delta. \quad (30)$$

If $|h_n| > \Delta$, then

$$|h_n^*| < |h_n| \quad (31)$$

and each output sample that we desire to zero is always decreased.

On the other hand, if $|h_n| < \Delta$, then

$$|h_n^*| < 2\Delta. \quad (32)$$

Thus the process must converge to within an error of 2Δ on each output sample. As the step size Δ goes to zero the truncated distortion D_N approaches zero, which implies that the over-all distortion D has been minimized.

3.2 Equalizer Implementation

An equalizer implementing this strategy is shown in Fig. 4. This equalizer employs a 13-tap delay line with 6 variable taps on either side of the reference tap. The action of the equalizer is as follows: A succession of test pulses is sent through the transmission line and transversal filter. As each test pulse comes out of the transversal filter it is sliced to retain only polarity information and then sampled at T -second intervals. These polarity samples are stored in a 12-stage shift register. When the shift register is full, a gate is opened, and all taps are simultaneously adjusted one step up or down in attenuation according to the polarities in the shift register. At this time a pulse height or normalization adjustment is also made on the over-all gain by means of the upper slicer-sampler circuit shown in Fig. 4. The method of obtaining electronically controlled steps of attenuation for the equalizer taps is shown in Fig. 5. It can be seen that reversible counters, directed by the 12-stage shift register, control the attenuation of a ladder network by means of relays.

In the particular equalizer which has been constructed, the quantum steps on the tap gain controls are about 0.25 per cent ($\Delta = 0.0025$). For high-quality voice channels utilized by the VSB data system, about 100 test pulses are required for settling of the tap attenuators to optimum values. This settling time depends on the size of the quantum step, the initial distortion, and the channel noise. The settling time and residual distortion are the subject of the following section on

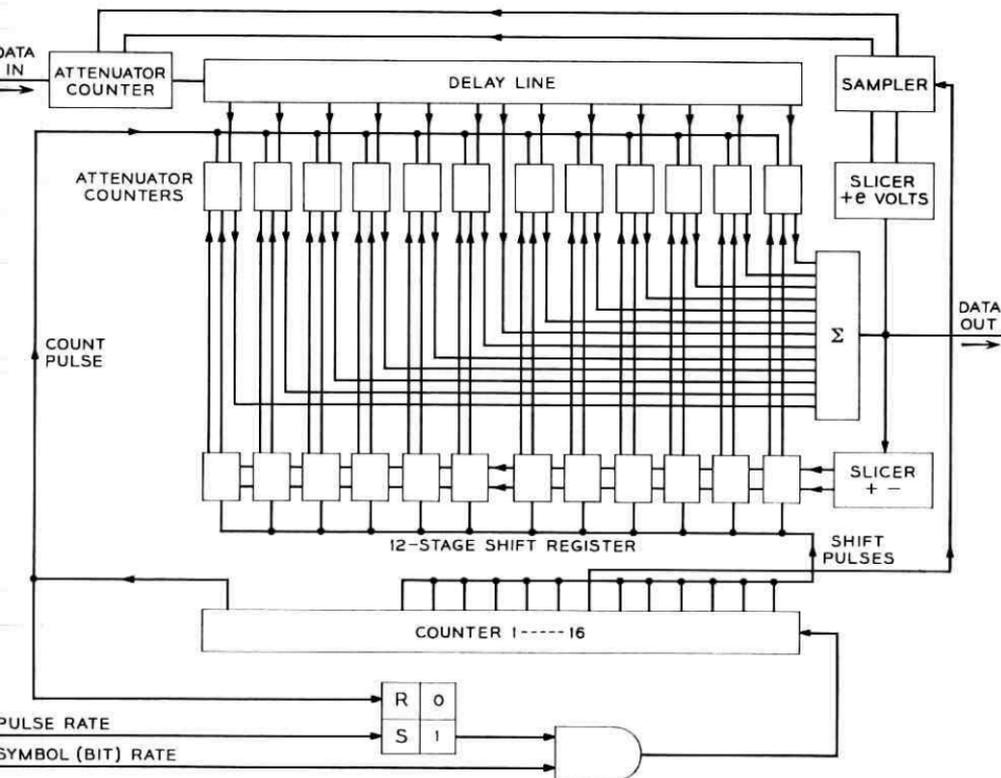


Fig. 4 — Automatic equalizer.

equalizer performance. Note that the equalizer is relatively unaffected by large impulse noise, since the output voltage is sliced and a large impulse can cause only one wrong step, which is subsequently corrected.

Figs. 6 and 7 show the results of equalization in a typical example using the 12-tap automatic equalizer. The single-pulse response together with the binary and 8-level eye patterns are shown before and after equalization. Before equalization only binary transmission was possible in this example. Since the equalizer has enabled 8-level operation in place of binary, a threefold increase in speed capability of the channel has been obtained.

3.3 The Initial Distortion Limitation

Thus far we have only considered equalization strategy when the initial distortion is less than 100 per cent. When $L_0 > 1$, there are two

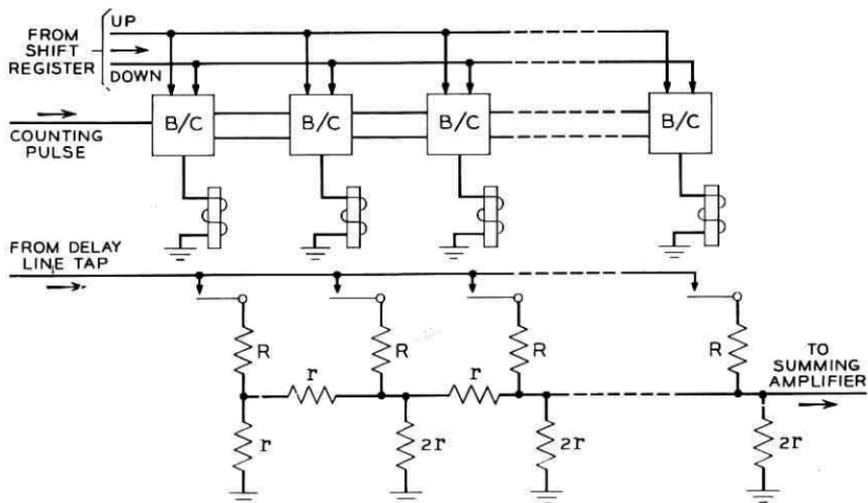


Fig. 5 — Reversible counter-controlled attenuator.

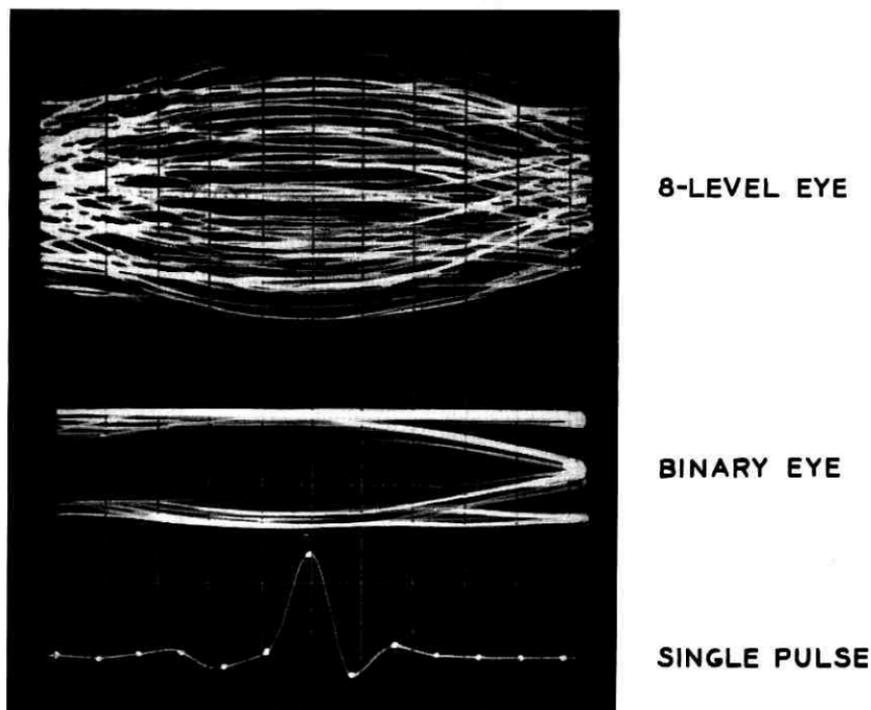


Fig. 6 — Unequalized pulse and corresponding "eye patterns."

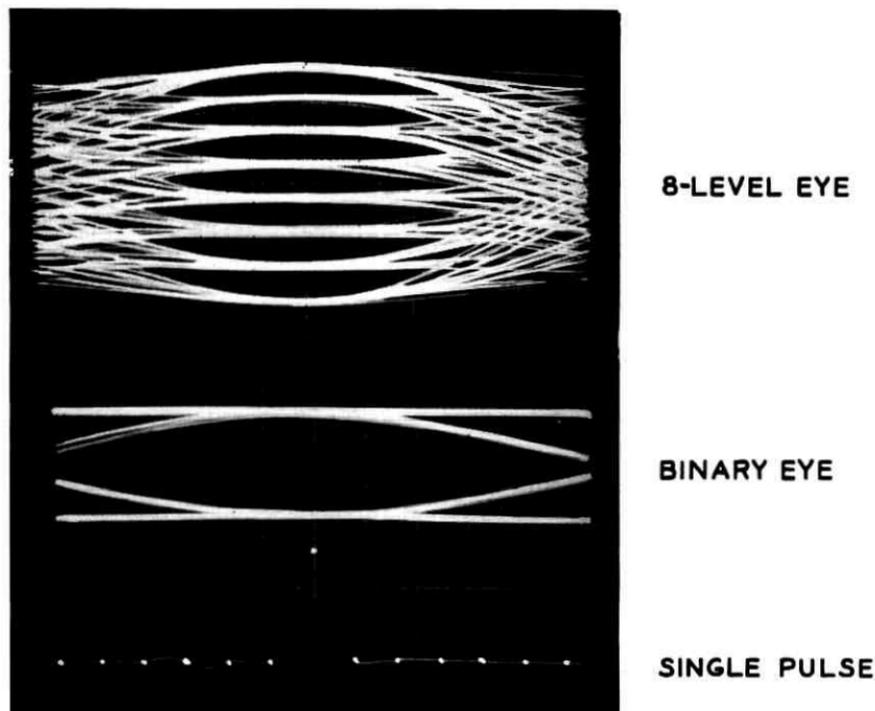


Fig. 7—Equalized pulse and corresponding “eye patterns.”

sources of failure for this equalizer. First, the equalizer convergence algorithm may fail to converge to the tap gains which force zeros in the output response for $t = -(N/2)T, \dots, (N/2)T$. Second, these settings may not be the optimum (minimum distortion) gains. In either case the condition $D_0 < 1$ is *sufficient* but not *necessary* to prevent failure, and it is quite possible that the equalizer will converge and be optimum over a wider range of inputs.

In computer simulations of the equalizer strategy it has thus far been impossible to induce failure of the convergence algorithm without causing failure in optimality. Examples of the converse are, however, easily constructed. One such example is illustrated in Figs. 8 and 9. The channel to be equalized in this example has a 50 per cent cosine roll-off amplitude characteristic and a linear delay characteristic. The final distortion after equalization with an N -tap equalizer is shown in Fig. 8 for two different values of peak delay. With a peak delay of 5 pulse intervals the initial distortion $D_0 = 1.4$. In this case it happens

that the equalizer converges and is optimum. The more taps N which are used on the equalizer, the lower the final distortion.

When the peak delay in this example is increased to 6 pulse intervals, we encounter a radically different behavior. The equalizer still converges to the tap settings which satisfy $h_n = 0$ for $|n| \leq N/2$, $n \neq 0$, but this solution no longer minimizes distortion. The initial distortion here is about 2.1 and after equalization with a 6-tap equalizer the final distortion is over 3. Clearly this is not optimum, since the zero tap settings gave better performance. As is shown, increasing the number of taps results in larger residual distortion.

The equalizer's failure in this latter case is more clearly illustrated in Fig. 9. The impulse response before equalization looks harmless enough, but after the equalizer has forced 3 zeros on either side of the response peak (6-tap equalizer) a large side lobe is created outside the equalizer time range on the right-hand side of the response. This side lobe contains more distortion than was present in the original response. In this

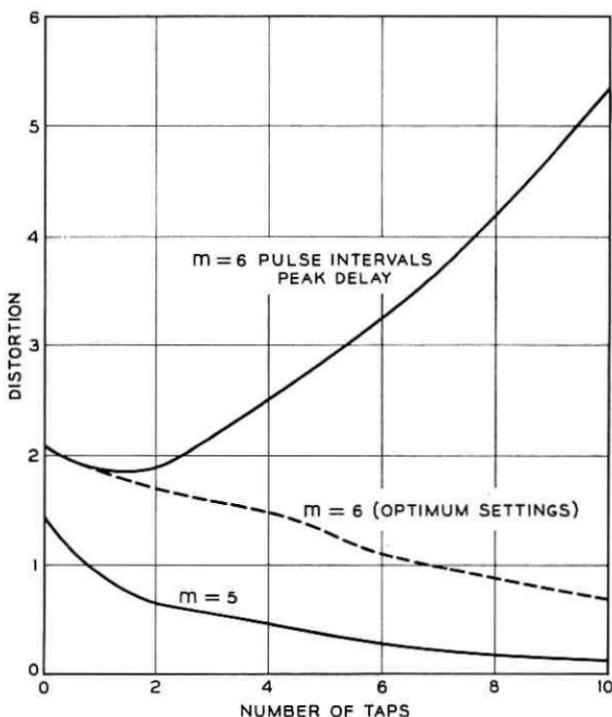


Fig. 8—Example of equalizer failure—linear delay.

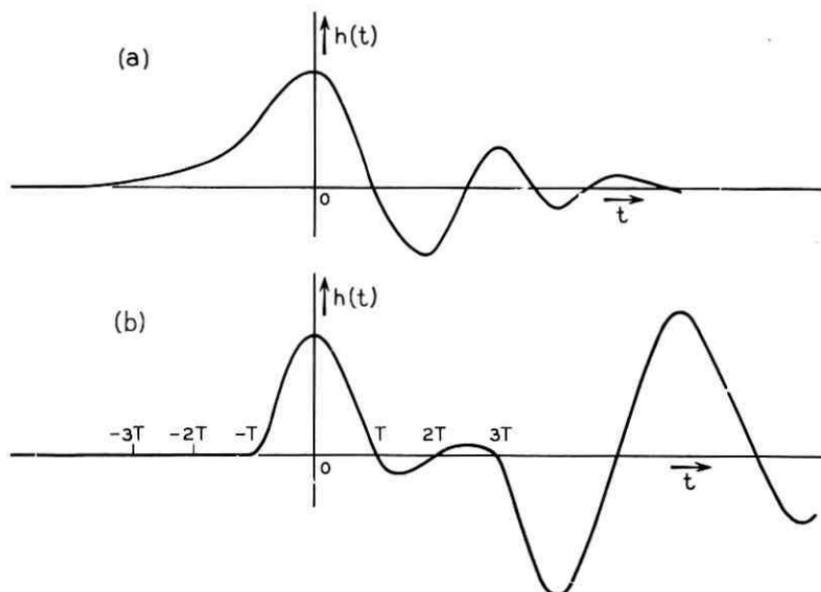


Fig. 9—(a) Impulse response for channel with linear delay, $m = 6$; (b) after equalization with 6-tap equalizer.

example minimum distortion settings of the equalizer do not set zeros at the samples h_n for $|n| \leq (N/2)$, $n \neq 0$. The actual optimum settings may be obtained by methods we will discuss presently, and for comparison purposes the minimum obtainable distortion for this example is plotted in Fig. 8.

3.4 Strategy When $D_0 > 1$

We described the situation when the initial distortion was less than 100 per cent as "loosely coupled." In this range many iterative schemes can be devised which converge to the optimum. Conversely, when $D_0 > 1$ the tap gains become more strongly interdependent. Also, the optimum may shift so that it is impossible to simply instruct the automatic equalizer to force N zeros in the impulse response at the locations of the N variable taps of the equalizer. Thus the equalization strategy for the general case becomes more subtle and complicated than the simple strategy previously explained.

Fortunately, Theorem II describes D (using center-tap normalization) as a convex function of the tap gains. Therefore the distortion has a single minimum, and mathematical programming methods may be

used to locate the optimum tap gains. A steepest-descent method may be applied to D [see (14)]. The gradient of D is

$$\bar{\nabla}D = \sum_{j=-N/2}^{N/2} \frac{\partial D}{\partial c_j} \bar{a}_j \quad (33)$$

$$\frac{\partial D}{\partial c_j} = \sum_{n=-\infty}^{\infty} (x_{n-j} - x_n x_{-j}) \operatorname{sgn} h_n. \quad (34)$$

If each tap gain c_j is incremented proportional to its gradient component (34), the system will eventually approach the optimum tap gains. The dotted lines showing minimum distortion in Fig. 8 were calculated using this strategy on a digital computer.* Such a strategy can be implemented with a modification of the circuitry previously described. In this implementation the term involving $x_n x_{-j}$ in (34) is neglected as small in comparison to x_{n-j} . The remainder of (34) may be derived physically by a two-pulse cycle. The first test pulse is sliced to obtain the sequence $\{\operatorname{sgn} h_n\}$ which is stored in the shift register. This sequence is then used to control the polarity of the gains at each tap. A second test pulse is transmitted and the transversal filter serves to multiply the input sequence $\{x_n\}$ by the tap gain sequence $\{\operatorname{sgn} h_n\}$. The output voltage is approximately the sequence $\{\partial D/\partial c_j\}$ from (34) and may be used to digitally increment the tap attenuators.

In any such scheme it is now necessary to have a transversal filter about twice as long as the number of variable taps to be used. This is because the sum in (34) is infinite, but practically speaking an N -tap transversal filter will affect the impulse response for not more than $2NT$ seconds. Thus the test pulses must go through a $2N$ -tap delay line, all of whose taps are equipped to handle the ± 1 gains which store the sequence $\{\operatorname{sgn} h_n\}$. However, only the inner N taps need have associated variable attenuators. This complication comes about because the distortion depends on what happens outside the NT range (witness Fig. 9!) as well as what happens inside. This information must be measured by "listening posts" established by taps on the delay line outside the normal range and then taken into consideration in incrementing the variable tap gains.

IV. THEORETICAL PERFORMANCE OF THE EQUALIZER

4.1 Design and Performance Parameters

Our attention will now be confined to the equalizer described in Section 3.2 and illustrated in Fig. 4. In this fairly simple system there

* The distortion minimization can also be formulated as a linear programming problem, so that using a digital computer the exact optimum can be reached in a finite number of steps.

are only two parameters which the designer has under his control to affect the performance and cost of the system. These are:

N — the number of variable taps on the transversal filter.

Δ — the step spacing on each electronically controlled attenuator. (More generally this is Δ_f , since different tap positions may take different-sized steps.)

The economics of the choice of these two parameters may be readily appreciated. The cost of the equalizer is nearly directly proportional to N , since not only does a larger N entail a proportionally longer delay line, but also proportionally more logic circuitry. A major portion of this logic circuitry is taken by the reversible counters for setting the attenuator tap coefficients. To decrease the step spacing Δ , each of the counters must be augmented by additional stages.

In judging the performance of the equalizer we shall be interested in the following two parameters;

D_r — the final distortion or residual distortion after equalization.

T_s — the settling time or the time required to set the equalizer during the training period.

Obviously, the smaller the residual distortion D_r , the better the data system will perform, but on the other hand it is too expensive to attempt to reduce this residual distortion much below the required tolerance for a given system. The settling time T_s on an experimental implementation of the automatic equalizer has been on the order of a second. When compared with the time required for establishing the call and acquiring timing synchronization this seems negligible. However, we shall find that when greater accuracy is required the time T_s can become quite appreciable.

The performance parameters D_r and T_s depend on the design parameters N and Δ and upon the channel to be equalized. Since it has been common to describe the channel by frequency-domain characteristics we shall do so here. The channel characteristics of interest are:

$A(\omega)$ — the amplitude characteristic of the system, including transmitter shaping filter and the attenuation characteristic of the channel.

$\beta'(\omega)$ — the delay characteristic of the channel.

S/N — the signal-to-noise ratio of the channel. The noise is assumed to be Gaussian and white since the effects of impulse noise are not particularly important in the automatic equalization system.

Fig. 10 shows how the parameters N and Δ affect the performance of a noiseless system example. In these curves the distortion is plotted as a

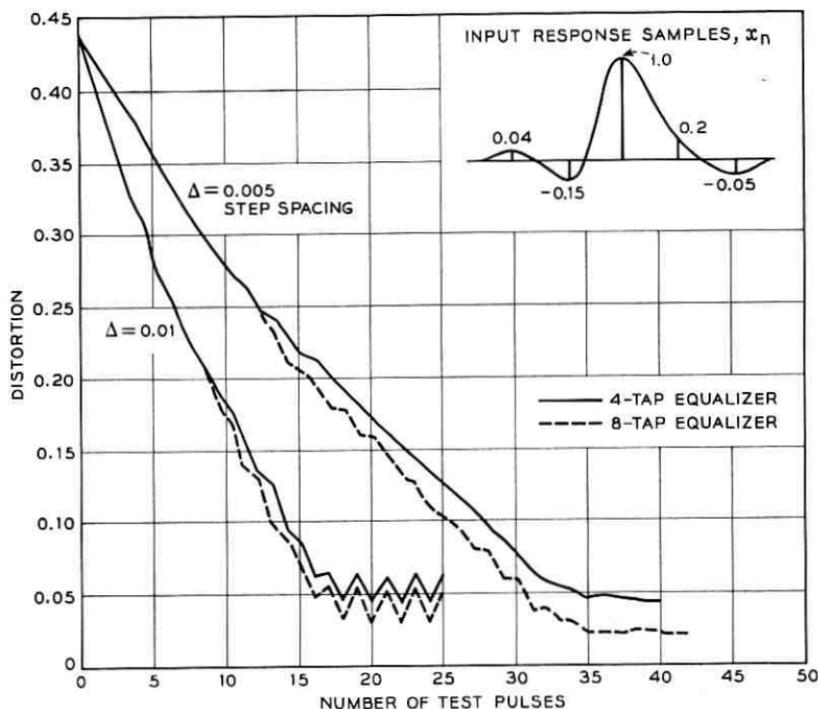


Fig. 10—Distortion vs time.

function of time. It can be seen that increasing the number of taps N results in a lower residual distortion without affecting the settling time of the system. After the required settling time has elapsed, the system reaches a limit cycle where all taps oscillate one step up and down about the optimum values. Decreasing the step spacing Δ decreases the oscillation of distortion and increases the settling time required. When noise is introduced into the model, the curve of distortion versus time becomes a random walk with final values of distortion becoming random variables.

The residual distortion is composed basically of two components. One component is the theoretical minimum distortion for the given channel and length N of transversal filter available. This corresponds to the case where $S/N \rightarrow \infty$ and $\Delta \rightarrow 0$. No settings of the filter can reduce the residual distortion any further. This distortion component will be designated D_c . It is a function only of the channel characteristics $A(\omega)$ and $\beta'(\omega)$ and of the number of taps N .

The second component of residual distortion arises from our inability

to reach optimum tap settings because of the oscillation of the final tap settings due to noise and finite step spacing Δ . This component is approximately independent from tap to tap. On a per-tap basis it is a function only of step spacing Δ_j and signal-to-noise ratio. This component will be designated D_s , where D_s is the contribution from a single tap. Thus for an N -tap equalizer the distortion may be resolved as follows.

$$D_r = D_c(N, A, \beta') + \sum_{j=-N/2}^{N/2} D_s(\Delta_j, S/N). \quad (35)$$

The subscripts "c" and "s" stand for channel distortion and system distortion respectively.

In the next section we will discuss generally D_c , the channel distortion, and show curves relating D_c and N for various shapes of amplitude and delay distortion. The subsequent section then deals with the system distortion D_s . Curves are shown relating D_s and Δ for various signal-to-noise ratios. The question of settling time is also discussed in this latter section.

4.2 Residual Channel Distortion

When $D_0 < 1$, the minimum distortion is obtained by setting $h_n = 0$ for $|n| \leq N/2$, $n \neq 0$. The remaining distortion we describe as D_c , the residual channel distortion.

$$D_c = \frac{1}{h_0} \sum_{|n| > N/2} |h_n|. \quad (36)$$

This expression for the residual channel distortion is easily calculated on a digital computer from N , $A(\omega)$, and $\beta'(\omega)$ by first computing the response samples $\{x_n\}$ from the Fourier transform of the channel's frequency characteristic and then solving the N simultaneous equations $h_n = 0$ for $|n| \leq N/2$, $n \neq 0$. A number of curves obtained by this procedure are presented in Figs. 11 through 13. In all of these figures the transmitter shaping characteristic was raised cosine with a 50 per cent roll-off. (Full raised cosine shaping gives very similar results.) The reference time was taken at the peak of the response $x(t)$. This is not necessarily optimum, but usually is fairly close to optimum.

Fig. 11 shows the residual channel distortion for parabolic delay over a wide range of equalizer size N . Fig. 12 is a similar presentation of the residual distortion resulting from linear delay. In Fig. 13 the channel has no delay distortion, but has an attenuation characteristic which

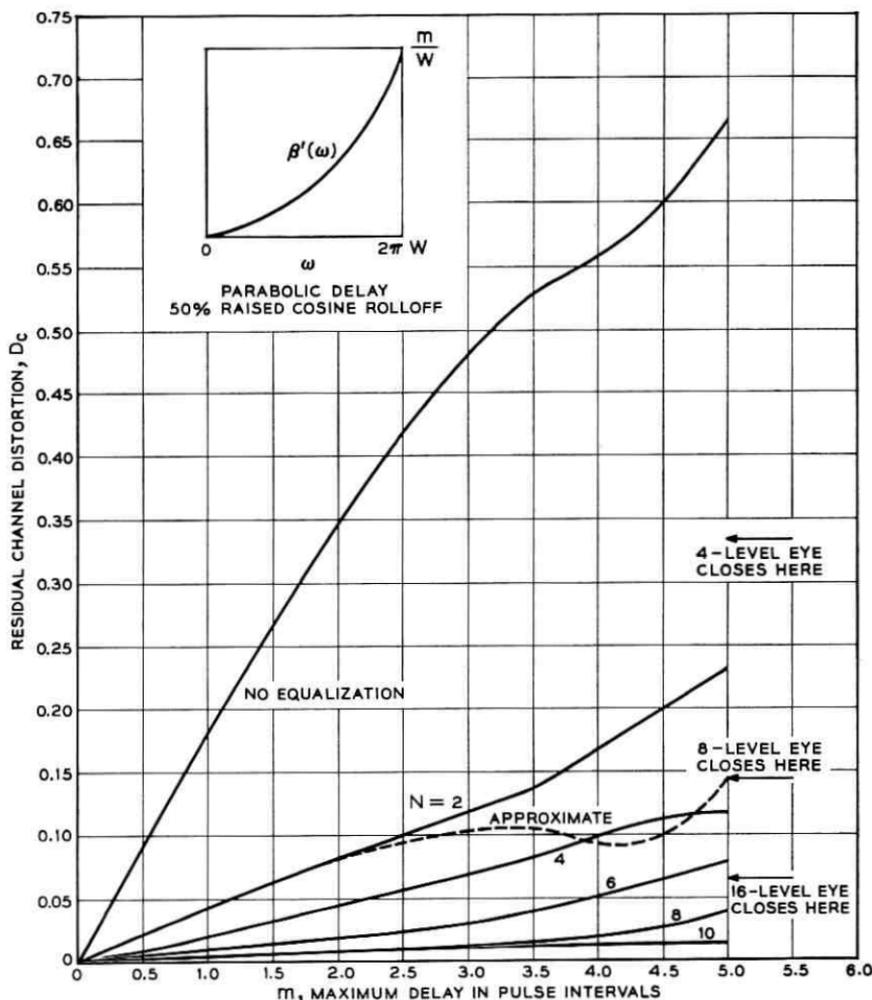


Fig. 11—Residual channel distortion for an N -tap equalizer with parabolic delay.

falls off with various slopes. These curves are intended to answer the question of how well a given transversal filter can equalize a given type of distortion. These curves represent theoretical minima. To see how well an actual equalizer can approach these values, the system distortion D_s for the given Δ and S/N must be added. From (36) the residual distortion consists of all the response samples h_n removed from the center of the response by more than $nT/2$ seconds. This is approxi-

mately composed of two components: what distortion was already out there in the original channel response $x(t)$, and what distortion has been squeezed out into this range by the equalization process. We know from echo theory that the former component results from delay and amplitude ripples of greater than $N/2$ cycles in the bandwidth and increases approximately linearly with the amplitude of the ripple content. The latter component of the residual channel distortion which is squeezed outside the transversal filter range is more difficult to conceptually visualize.

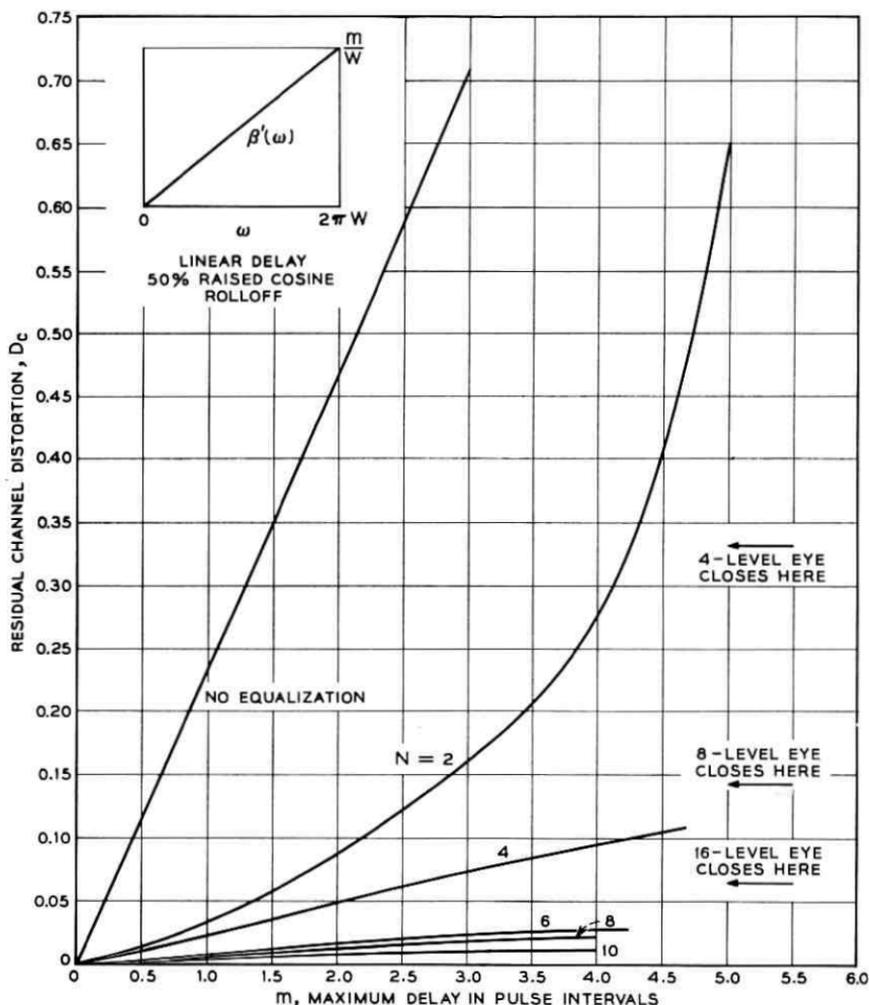


Fig. 12 — Residual channel distortion for an N -tap equalizer with linear delay.

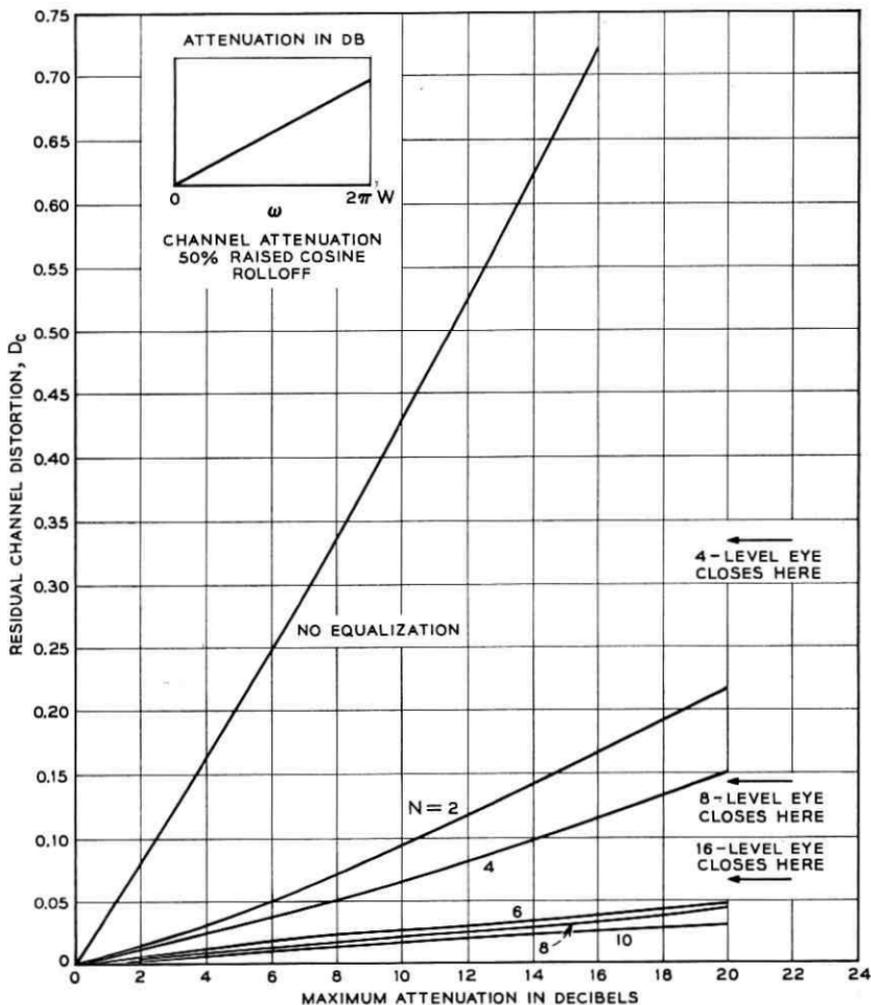


Fig. 13—Residual channel distortion for an N -tap equalizer with attenuation distortion present in the transmission channel.

We can show that this component of D_c increases approximately with the square of the delay or amplitude variation in the channel.

Removing the large terms involving c_0 and x_0 , which are unity, from under the summation in (9) we arrive at

$$h_n = x_n + c_n + \sum_{\substack{j=-N/2 \\ j \neq 0, n}}^{N/2} c_j x_{n-j}. \quad (37)$$

In order to make $h_n = 0$ for $-N/2 \leq n \leq N/2$, $n \neq 0$, it can be seen that a first approximation for the tap setting c_j is

$$c_j \approx -x_j. \quad (38)$$

After equalization we will have $h_n = 0$ inside the range of the filter. Outside the range we will have approximately, using (37) and (38)

$$h_n \approx x_n - \sum_{\substack{j=-N/2 \\ j \neq 0, n}}^{N/2} x_j x_{n-j}. \quad (39)$$

The first term in (39) is the initial distortion term and the second is distortion outside the range of the filter generated by the equalization process. Since the samples x_n vary approximately linearly with amplitude and delay ripple amplitude, it can be seen that the two components of h_n vary linearly and quadratically with ripple amplitude respectively.

The curves in Figs. 11 through 13 show that the linear term dominates in the case of parabolic or linear delay in the channel. This indicates that very little additional distortion in the range $|t| > NT/2$ is being pushed out by the equalization process. In Fig. 11 directly beneath the $N = 2$ curve is plotted the initial distortion outside the range of this particular equalizer. This dotted curve marked "approximate" indicates the quantity

$$D_r(\text{approx.}) = \sum_{|n| > 1} |x_n|. \quad (40)$$

In the upper range of the $N = 2$ curve the second component involving "squeezed-out" distortion becomes important.

As long as the linear term dominates we can think of the transversal filter as completely equalizing any delay or amplitude variation of less than $N/2$ cycles in frequency content. For example, a Fourier expansion of the parabolic delay reveals the major content at low frequencies; thus few taps are required to equalize parabolic delay. The higher-frequency content of the parabola is unequalizable by a short transversal filter.

Frequently the delay and amplitude characteristic of a channel consist of approximately sinusoidal ripples. This is generally the case when an equalization network has been incorporated in the channel. A good rule of thumb is that it takes twice as many taps on the equalizer as there are ripples in the bandwidth. Thus the better a channel has been previously equalized the longer a transversal filter will be required to further improve equalization. Frequency- and time-domain equaliza-

tion are very similar in the respect that they both leave ripples of delay and amplitude of higher frequency after equalization.

4.3 Residual System Distortion

The term D_s , residual system distortion, results from the inability of the equalizer to reach the actual optimum tap settings. In the case of very high signal-to-noise ratio the value of D_s is obvious. The tap c_j will end in a limit cycle of one step Δ_j about the optimum tap setting. In one position

$$c_j = c_j(\text{optimum}) + \epsilon \quad (41)$$

while in the other position

$$c_j = c_j(\text{optimum}) + \epsilon - \Delta_j. \quad (42)$$

Assuming that $\Delta_j > |\epsilon|$ and neglecting the last term in (37), we obtain the average system distortion for this case.

$$D_s(\Delta_j, (S/N) = \infty) = \frac{1}{2}\Delta_j. \quad (43)$$

Thus the system distortion is one-half step for high S/N . However, a system will very seldom operate in a true high- S/N environment because the tap steps themselves are usually well within the noise. For example, with a Δ of 0.01 the step-to-noise ratio is 40 db below the signal-to-noise ratio. With noise the taps end in a random walk instead of the limit cycle of the noiseless case, and the average system distortion will be considerably higher than that given by (43).

Let's examine the behavior of an individual sample h_j when noise is present in the channel. The receiver then bases its decision on whether to advance or retard tap c_j by the amount Δ on the sign of the quantity $(h_j + n_j)$, where n_j is Gaussian with mean zero and variance σ^2 , and is independent over index j and from test pulse to test pulse. Assuming that the sample h_j is affected only by tap setting c_j [equivalent to dropping the last term in (37)] we write

$$h_j \approx x_j + c_j. \quad (44)$$

Since c_j takes on only values which are integral multiples of Δ , we can without any particular loss of generality quantize h_j in steps of Δ . The behavior of each of the samples h_j is similar, so in order to keep confusion to a minimum we shall drop the "j" index and define

$$\text{Prob } [h = k\Delta, \text{ after } m \text{ test pulses}] = p(k, m). \quad (45)$$

Now we can write a difference equation for this probability as follows

$$p(k, m+1) = p(k-1, m)P(k-1) + p(k+1, m)Q(k+1) \quad (46)$$

where $P(k)$ and $Q(k)$ are the probabilities of noise being less than or greater than $-k\Delta$ respectively.

$$P(k) = \int_{-\infty}^{-k\Delta} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \frac{1}{2}[1 - \text{Erf}(k\Delta/\sqrt{2}\sigma)] \quad (47)$$

$$Q(k) = 1 - P(k). \quad (48)$$

Initially the value h_j starts at position $x_j \approx l\Delta$ so that we can use the initial condition

$$p(k, 0) = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l. \end{cases} \quad (49)$$

The difference equation (46) then defines a probability distribution which spreads out as time (m) progresses and eventually ends in a stable symmetrical distribution centered at zero. [Actually there are two final distributions reached: one for m even and one for m odd. Since the system could be turned off in either state with equal likelihood, these distributions must be averaged. This effect is equivalent to the averaging of (41) and (42) in the noiseless case.]

A number of these final distributions were computed using (46), (47) and (48) on a digital computer. The contribution to the residual distortion owing to oscillation of tap c_j is the random variable $|h_j|$. The system distortion $D_s(\Delta_j, S/N)$ is defined as the expected value of $|h_j|$

$$D_s = E[|h_j|] = \Delta_j \sum_k |k| p(k, \infty). \quad (50)$$

A number of curves relating D_s , Δ and signal-to-noise ratio are shown in Fig. 14. Observe that these curves are nearly piecewise linear with a break point approximately where $\Delta = \sigma$. Above this point the high signal-to-noise condition prevails and $D_s \approx \frac{1}{2}\Delta$. Below this point the steps Δ are within the noise and the slope of the curve changes. By making a low signal-to-noise approximation we can derive equations for these curves.

For Δ/σ small we use the first term of the series

$$\text{Erf}(X) = \frac{2X}{\sqrt{\pi}} \left[1 - \frac{X^2}{1!3} + \frac{X^4}{2!5} - \frac{X^6}{3!7} + \dots \right] \quad (51)$$

and obtain

$$P(k) = \frac{1}{2} - \frac{k\Delta}{\sqrt{2\pi}\sigma}. \quad (52)$$

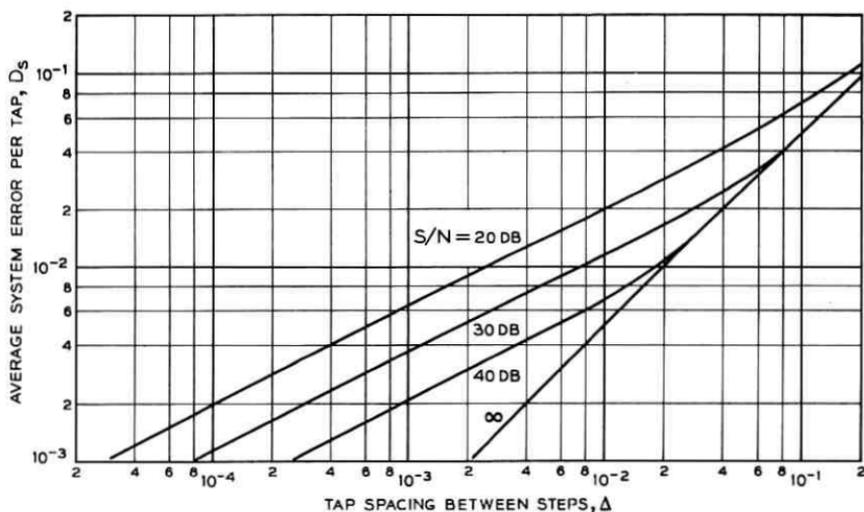


Fig. 14— D_s vs Δ for various S/N ratios.

Using this transition probability in the difference equation (46) we obtain an equation similar to the classical problem of the random walk of an elastically bound particle:

$$p(k, m+1) = p(k-1, m) \left[\frac{1}{2} - \frac{(k-1)\Delta}{\sqrt{2\pi}\sigma} \right] + p(k+1, m) \left[\frac{1}{2} + \frac{(k+1)\Delta}{\sqrt{2\pi}\sigma} \right]. \quad (53)$$

An exact solution to this difference equation assuming the initial condition (49) is given by M. Kac in Ref. 3. Since this solution is more complicated than the asymptotic distribution obtained as $\Delta \rightarrow 0$ and is no more useful for our purposes, we will not repeat Kac's formula here.

It is simple to derive expressions for the mean and variance of h_j after m test pulses have been transmitted. Using (53) we obtain for the mean value

$$E[h(m)] = \Delta \sum_k k p(k, m) \quad (54)$$

$$E[h(m)] = \Delta \left[\frac{1}{2} + \frac{\Delta}{\sqrt{2\pi}\sigma} \right] \sum_k \cdot k [p(k-1, m-1) + p(k+1, m-1)] + \frac{\Delta^2}{\sqrt{2\pi}\sigma} \sum_k k^2 [p(k+1, m-1) - p(k-1, m-1)]. \quad (55)$$

The various summations in (55) can be rearranged with changes in variable to give

$$E[h(m)] = E[h(m-1)] \left(1 - \frac{2\Delta}{\sqrt{2\pi}\sigma} \right). \quad (56)$$

Since the initial value of h_j is $x_j \approx l\Delta_j$, the average value of h_j after m test pulses is

$$E[h_j(m)] = l\Delta_j \left(1 - \frac{2\Delta_j}{\sqrt{2\pi}\sigma} \right)^m. \quad (57)$$

One can derive a similar expression for the variance of h_j after m test pulses

$$E[h_j^2(m)] = (l\Delta_j)^2 \left(1 - \frac{4\Delta_j}{\sqrt{2\pi}\sigma} \right)^m + \frac{\sqrt{2\pi}}{4} \Delta_j \sigma \left[1 - \left(1 - \frac{4\Delta_j}{\sqrt{2\pi}\sigma} \right)^m \right]. \quad (58)$$

As $m \rightarrow \infty$ the final distribution is obtained and

$$\lim_{m \rightarrow \infty} E[h_j(m)] = 0 \quad (59)$$

$$\lim_{m \rightarrow \infty} E[h_j^2(m)] = \frac{\sqrt{2\pi}}{4} \Delta_j \sigma. \quad (60)$$

Equation (60) is particularly interesting, since the variance of h_j is proportional to the square root of the noise variance σ^2 .

Now suppose the tap-to-noise ratio Δ/σ tends to zero and test pulses are sent at a rate $1/\tau$ which tends to infinity. Using the difference equation (53) we can derive a differential equation for the probability density of the sample h_j after t seconds have elapsed:

$$\frac{\partial p(h,t)}{\partial t} = \frac{2\Delta}{\sqrt{2\pi}\sigma\tau} \left[p(h,t) + h \frac{\partial p(h,t)}{\partial h} \right] + \frac{\Delta^2}{2\tau} \frac{\partial^2 p(h,t)}{\partial h^2}. \quad (61)$$

A solution to this equation, attributed to Lord Rayleigh, is given in Ref. 3. Assuming h_j starts at x_j , the probability density after t seconds is

$$p(h_j, t) = \frac{1}{\sqrt{2\pi}\alpha(1 - e^{-2\gamma t})} \exp [-(h_j - x_j e^{-\gamma t})^2 / 2\alpha^2(1 - e^{-2\gamma t})], \quad (62)$$

where α^2 is the same as $E[h_j^2(\infty)]$ in (60)

$$\alpha^2 = \frac{\sqrt{2\pi}}{4} \Delta_j \sigma \quad (63)$$

and

$$\gamma = \frac{2\Delta_j}{\sqrt{2\pi} \sigma_T}. \quad (64)$$

As $t \rightarrow \infty$ the density (62) tends to Gaussian of mean zero and variance α^2 . We can thus easily derive the low signal-to-noise ratio value of D_s .

$$D_s = E[|h_j|] = 2 \int_0^\infty \frac{h_j}{\sqrt{2\pi} \alpha} \exp - (h_j^2/2\alpha^2) dh_j \quad (65)$$

$$D_s = \sqrt{2/\pi} \alpha = (\sigma\Delta_j/\sqrt{2\pi})^{1/2} \quad (66)$$

$$D_s \approx 0.633 \sqrt{\sigma\Delta_j}. \quad (67)$$

The straight-line portions of the curves D_s versus Δ in Fig. 14 match nearly exactly with (67).

In summary, we have found

$$D_s \approx \begin{cases} \Delta_j/2 & \text{if } \Delta_j > \sigma \\ 0.633\sqrt{\sigma\Delta_j} & \text{if } \Delta_j < \sigma. \end{cases} \quad (68)$$

The important fact about (68) is that the residual system error goes down as the square root of both tap spacing Δ and the standard deviation of the noise σ . Thus once the step spacing and noise are comparable, cutting the steps Δ by a factor of ten cuts the residual system error due to tap uncertainty by about 3.16. An interesting sidelight is that cutting Δ by ten might seem equivalent to averaging ten samples of h_j before taking one step of Δ . In the latter case the noise σ would be cut by $\sqrt{10} = 3.16$ and the residual system distortion D_s cut by only $\sqrt{3.16} = 1.77$. However, as might be guessed this latter technique will require less settling time than the former technique using 0.1Δ and moving ten times as fast.

This leads us into an abbreviated discussion of settling time. The time required to settle is usually determined by the largest distorting sample x_j for $j \neq 0$. If the step-to-noise ratio Δ/σ is large, then the settling time is obviously

$$T_s = x_{j\max}/\Delta_j. \quad (69)$$

When the noise becomes important, then we need to make some sort of arbitrary definition of settling time. This definition could be based on the time required for the mean of h_j or the variance of h_j to reach some predetermined position or percentage, or it could be based on the time

constant $1/\gamma$ in the density given by (62). All these types of definition lead to similar expressions differing chiefly by constants, so we'll use the simplest and define

$$T_s = 1/\gamma = \sqrt{2\pi} \sigma\tau/2\Delta_j. \quad (70)$$

Recall that τ is the interval between test pulses. This expression is independent of $x_{j\max}$ for low tap-to-noise ratios. Using a binomial expansion of the term to the m th power in (57), (70) can be derived as the point where the first two terms of the expansion for $E[h_j(m)]$ cancel.

Going back to our previous example, where we compared decreasing the step spacing Δ with averaging several samples of h_j , we can now show that cutting the tap spacing Δ by a factor of N is exactly equivalent in settling time and system error to averaging N^2 consecutive samples of h_j . This is evident from the proportionalities involved in (69) and (70). In either case the same amount of time is required to achieve a given level of accuracy in equalization.

V. FREQUENCY-DOMAIN CONSIDERATIONS

5.1 *Frequency-Domain Relationship*

The equalizer that has been described is intended for the correction of distortion in digital data transmission. This equalizer is strictly a time-domain device which corrects the pulse response of the channel to the best ability of a finite-length tapped delay line. The question which most frequently arises concerning its operation asks what happens to the frequency characteristics of the channel as a result of the time-domain equalization. This question is asked not only out of curiosity and because engineers tend to think in terms of the frequency domain, but also because the time-domain equalizer is sometimes considered for the equalization of analog channels.

In this section we will develop a general formula for the frequency-domain characteristics of an equalized channel in terms of the un-equalized characteristics. This formula allows us to compute final characteristics and to make a few general observations about the relationship between initial and final characteristics. In addition to these results we are able to derive conditions as to when a channel may be equalized successfully, and we find one additional technique of computing optimum tap settings for long equalizers.

Henceforth we will assume an infinite-length transversal filter. If the equalizer is long enough to do its job properly, the final characteristics will closely approximate the infinite length characteristics we will

derive here. An outline of this derivation is as follows. If the equalized pulse response is to have zero distortion, it must have Nyquist frequency characteristics implying symmetry in real and imaginary components. Knowing the channel characteristics and the type of characteristic capable of being assumed by the equalizer, we show there is only one such Nyquist characteristic the product can assume. This characteristic must be the final frequency response of the equalized channel.

The impulse response of the channel in terms of the amplitude, $A(\omega)$, and phase, $\beta(\omega)$, is

$$h(t) = \frac{1}{\pi} \int_0^{\infty} A(\omega) \cos [\omega t - \beta(\omega)] d\omega. \quad (71)$$

We assume that the time base has been adjusted by the removal of a flat delay (linear phase) component from $\beta(\omega)$ so that the peak of the output response occurs at time zero. For zero distortion we require

$$h_n = 0 = \frac{1}{\pi} \int_0^{\infty} A(\omega) \cos [n\omega T - \beta(\omega)] d\omega \quad \text{all } n, n \neq 0. \quad (72)$$

Changing (72) to use real and imaginary components we arrive at the equivalent condition

$$\int_0^{\infty} A_x(\omega) \cos n\omega T d\omega = 0 \quad \text{all } n, n \neq 0 \quad (73)$$

$$\int_0^{\infty} A_y(\omega) \sin n\omega T d\omega = 0 \quad \text{all } n \quad (74)$$

where

$$A_x(\omega) = A(\omega) \cos \beta(\omega) \quad (75)$$

$$A_y(\omega) = A(\omega) \sin \beta(\omega). \quad (76)$$

Let us assume that $A(\omega) = 0$ for $\omega > 2\pi/T$, i.e., the channel has no frequency component higher than twice the Nyquist band for its signaling rate. It would be an unusual case if this were not true. We now make a change in variables to shift the origin to the Nyquist frequency π/T and redefine the characteristics about this frequency.

$$\hat{A}_x(\omega) = A_x\left(\omega + \frac{\pi}{T}\right) \quad (77)$$

$$\hat{A}_y(\omega) = A_y\left(\omega + \frac{\pi}{T}\right). \quad (78)$$

(In general, a circumflex on any variable indicates that it is defined about the frequency π/T as origin.) With this change the conditions (73) and (74) become

$$\int_{-\pi/T}^{\pi/T} \hat{A}_x(\omega) \cos n\omega T d\omega = 0 \quad \text{all } n, n \neq 0 \quad (79)$$

$$\int_{-\pi/T}^{\pi/T} \hat{A}_y(\omega) \sin n\omega T d\omega = 0 \quad \text{all } n. \quad (80)$$

We also have the normalizing condition for the center sample.

$$h_0 = 1 = \frac{1}{\pi} \int_{-\pi/T}^{\pi/T} \hat{A}_x(\omega) d\omega. \quad (81)$$

Equation (79) says that $\hat{A}_x(\omega)$ must be orthogonal to all $\cos n\omega T$ between $-\pi/T$ and π/T . This implies that its Fourier expansion consists of a constant to satisfy (81) plus any arbitrary sine components. In other words it must be a constant plus any arbitrary odd function. Similarly, $\hat{A}_y(\omega)$ must be any arbitrary even function to satisfy condition (80). Simply stated, the conditions are then

$$\hat{A}_x(\omega) = (T/2) + \text{odd function} \quad (82)$$

$$\hat{A}_y(\omega) = \text{even function}. \quad (83)$$

These conditions were first derived by Nyquist,⁴ although engineers are generally more familiar with the case of flat delay where (82) and (83) reduce to the statement that $A(\omega)$ has Nyquist symmetry.

Now that we see the conditions that the product of the equalizer and channel responses must meet for perfect equalization, let's look at the type of response the equalizer alone can have. The Fourier transform of the equalizer's time response is

$$C(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{-jn\omega T} \quad (84)$$

where the c_n 's are the tap gain settings. As before we shift the origin of definition for $C(\omega)$ to π/T . The change in variables and definition gives

$$\hat{C}(\omega) = \hat{C}_x(\omega) + j\hat{C}_y(\omega) \quad (85)$$

with

$$\hat{C}_x(\omega) = c_0 + \sum_{n=1}^{\infty} (-1)^n (c_n + c_{-n}) \cos n\omega T \quad (86)$$

$$\hat{C}_y(\omega) = \sum_{n=1}^{\infty} (-1)^n (c_n - c_{-n}) \sin n\omega T. \quad (87)$$

Observe that $\hat{C}_x(\omega)$ is an even function and $\hat{C}_y(\omega)$ is an odd function.

Now let us suppose that the unequalized channel frequency response in terms of real and imaginary components defined about $\omega = \pi/T$ is

$$\hat{A}(\omega) = \hat{A}_x(\omega) + j\hat{A}_y(\omega). \quad (88)$$

The product of the unequalized frequency response $\hat{A}(\omega)$ from (88) and the equalizer response $\hat{C}(\omega)$ from (85) must satisfy the odd and even requirements on its real and imaginary components as given in (82) and (83). After separating the product $\hat{A}(\omega)\hat{C}(\omega)$ into real, imaginary, odd, and even components, it is possible to arrive at simultaneous equations for $\hat{C}_x(\omega)$ and $\hat{C}_y(\omega)$. The details of this process are quite straightforward and have been relegated to Appendix B.

The final equalizer response is found in terms of the even (labelled with an "e" subscript) and odd (labelled with an "o" subscript) components of the unequalized channel response. The equations derived are

$$\hat{C}_x(\omega) = \frac{\frac{T}{2} \hat{A}_{xe}(\omega)}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)} \quad (89)$$

$$\hat{C}_y(\omega) = \frac{-\frac{T}{2} \hat{A}_{yo}(\omega)}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)}, \quad (90)$$

where

$$\hat{A}_{xe}(\omega) = \frac{1}{2}[\hat{A}_x(\omega) + \hat{A}_x(-\omega)] \quad (91)$$

$$\hat{A}_{xo}(\omega) = \frac{1}{2}[\hat{A}_x(\omega) - \hat{A}_x(-\omega)]. \quad (92)$$

These equations determine the real and imaginary components of the equalizer frequency response.

Using this response for the equalizer, the equalized channel response becomes

$$\begin{aligned} \frac{2}{T} \hat{C}(\omega)\hat{A}(\omega) = & \left[1 + \frac{\hat{A}_{xe}(\omega)\hat{A}_{xo}(\omega) + \hat{A}_{yo}(\omega)\hat{A}_{ye}(\omega)}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)} \right] \\ & + j \left[\frac{\hat{A}_{xe}(\omega)\hat{A}_{ye}(\omega) - \hat{A}_{yo}(\omega)\hat{A}_{xo}(\omega)}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)} \right] \end{aligned} \quad (93)$$

which is the result we have been seeking.

5.2 Interpretation of the Frequency-Domain Relationship

The frequency response of the equalizer [from (89) and (90)] is

$$\hat{C}(\omega) = \frac{T/2}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)} [\hat{A}_{xe}(\omega) - j\hat{A}_{yo}(\omega)]. \quad (94)$$

Recall that $\hat{A}_{xe}(\omega)$ is the *even* component of the *real* part of the channel response taken about the Nyquist frequency, π/T . Similarly $\hat{A}_{yo}(\omega)$ is the *odd* component of the *imaginary* part of the response. Notice that $\hat{C}(\omega)$ does not depend on $\hat{A}_{xo}(\omega)$ or $\hat{A}_{ye}(\omega)$, so that these components may be specified arbitrarily without changing the equalizer settings. This points up the difference between frequency-domain equalization and time-domain equalization. Our data distortion criterion is such that we don't care about the components $\hat{A}_{xo}(\omega)$ and $\hat{A}_{ye}(\omega)$ of the frequency response. However, these components affect the shape of the final equalized characteristic, as can be seen from (93).

It is clear from (94) that the channel cannot be equalized if $\hat{A}_{xe}(\omega)$ and $\hat{A}_{yo}(\omega)$ are both zero for some ω , $0 \leq \omega \leq \pi/T$, in which case $\hat{C}(\omega)$ is unbounded. If this does not happen, practical considerations indicate that the channel is capable of being equalized. (These practical considerations require that both $[A(\omega)C(\omega)]$ and its Fourier transform are continuous and absolutely integrable.)

Equalization Condition:

A channel with complex gain $\hat{A}(\omega)$ is capable of being perfectly equalized with an infinite transversal filter if and only if the even component of the real part of $\hat{A}(\omega)$ and the odd component of the imaginary part of $\hat{A}(\omega)$ do not simultaneously vanish for some ω , $0 \leq \omega \leq \pi/T$.

Note that a channel can be equalized even if it transmits no energy in some interval $[\omega_1, \omega_2]$, $\omega_2 < \pi/T$, so long as the missing energy components are replaced in a symmetrically located region above π/T , i.e., $[(2\pi/T) - \omega_2, (2\pi/T) - \omega_1]$.

Now let's go back to (93) to try to get some feel for the shape of the final response. The equation itself is not complicated, but it unfortunately requires us to break $A(\omega)$ into even and odd components about the Nyquist frequency π/T and then further to break these components into real and imaginary parts. By this time we have almost no idea what the final response will be when we put everything back together as in (93). To take a special case which simplifies this process, suppose that the channel has perfectly flat delay, so all the "y" components are zero. Then

$$\hat{A}(\omega)\hat{C}(\omega) = \frac{T}{2} \left[1 + \frac{\hat{A}_{xo}(\omega)}{\hat{A}_{xe}(\omega)} \right]. \quad (95)$$

The first term represents a rectangular characteristic and the second term is an odd function about the frequency π/T . Thus we have a familiar Nyquist characteristic which is easily determined from the initial characteristic $A(\omega)$. Fig. 15 shows a sample case of this type having flat delay (linear phase). The components $\hat{A}_{xe}(\omega)$ and $\hat{A}_{xo}(\omega)$ and the equalized spectrum are shown for this example characteristic. With a little study it is possible to get a good feel for the sort of final characteristic which is obtained by equalization in this purely real case. However, the introduction of nonlinear phase results back in the complicated expression (93).

Typically, the channel to be equalized cuts off somewhere before $\omega = 2\pi/T$, so there are no energy components in the interval $[(\pi/T) + \omega_c, (2\pi/T)]$. In this case we have

$$\begin{aligned} \hat{A}_{xe}(\omega) &= -\hat{A}_{xo}(\omega) & \pi/T \geq \omega \geq \omega_c \\ \hat{A}_{ye}(\omega) &= -\hat{A}_{yo}(\omega) \end{aligned} \quad (96)$$

$$\begin{aligned} \hat{A}_{xe}(\omega) &= \hat{A}_{xo}(\omega) & -\pi/T \leq \omega \leq -\omega_c \\ \hat{A}_{ye}(\omega) &= \hat{A}_{yo}(\omega) \end{aligned} \quad (97)$$

Substitution of (96) and (97) into (93) shows that

$$\hat{A}(\omega)\hat{C}(\omega) = T/2 \quad \text{for} \quad -\pi/T \leq \omega \leq -\omega_c. \quad (98)$$

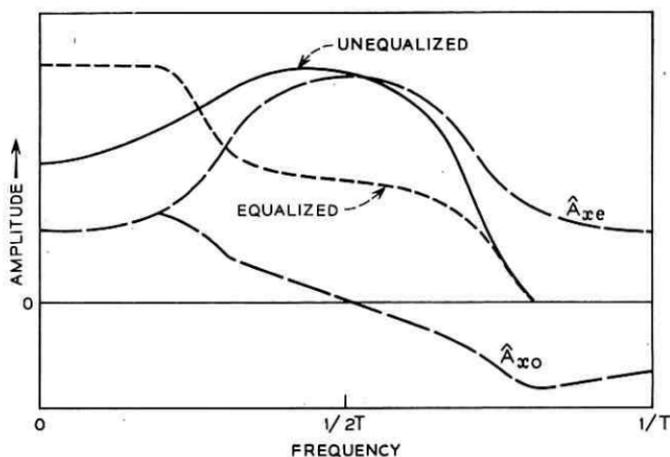


Fig. 15—Equalization of amplitude distortion.

Which means

$$A(\omega)C(\omega) = \frac{T}{2} \quad \text{for } 0 \leq \omega \leq \frac{\pi}{T} - \omega_c. \quad (99)$$

We can boil this equation down into the most significant observation we can make about the frequency-domain relationship:

If the channel response cuts off at the frequency $(\pi/T) + \omega_c$, then the channel must be equalized to constant amplitude and linear phase in the region $0 \leq \omega \leq (\pi/T) - \omega_c$.

This observation is important in the equalization of analog channels to flat amplitude and linear phase. In order to get the largest possible interval of perfect analog equalization we should arrange the tap spacing T so as to approach the Nyquist interval corresponding to the channel cutoff. However, if the channel cuts off *before* π/T then by the previous results the channel is unequalizable. The closer T is to a Nyquist interval the more taps will generally be required to effect a good equalization. One final note to add to the confusion — if ω_c is large very little of the spectrum will be equalized flat. If the tap spacing T cannot be changed we can increase this region of flat equalization by inserting a low-pass filter during the equalization period which has a cutoff frequency close above π/T . The filter may then be removed after the equalizer tap gains have been set.

Some of the principles involved are exemplified in Fig. 16, which shows the equalization of a channel which has a linear delay characteristic (quadratic phase). Initially the amplitude response of this channel is a 50 per cent raised cosine response cutting off at $\omega = 3\pi/2T$. This amplitude characteristic is a Nyquist shape in the absence of delay distortion. After equalization both amplitude and phase are flat from $\omega = 0$ to $\omega = \pi/2T$, as required by our previous observation. From $\omega = \pi/2T$ to $\omega = 3\pi/2T$ the phase still appears parabolic. In this region there has been an interaction between phase and amplitude which has resulted in a change in the amplitude characteristic away from its Nyquist shape. The combination of equalized phase and amplitude in this region is such as to satisfy conditions (82) and (83), although this is not evident from casual inspection.

One additional important usage of (93) for $\hat{C}(\omega)$ is in the calculation of optimum tap settings for the equalizer. Starting from the unequalized channel pulse time response samples $\{x_n\}$ we construct $\hat{A}_{x_0}(\omega)$ and $\hat{A}_{y_0}(\omega)$ exactly as in (86) and (87) for $\hat{C}_x(\omega)$ and $\hat{C}_y(\omega)$, i.e., let

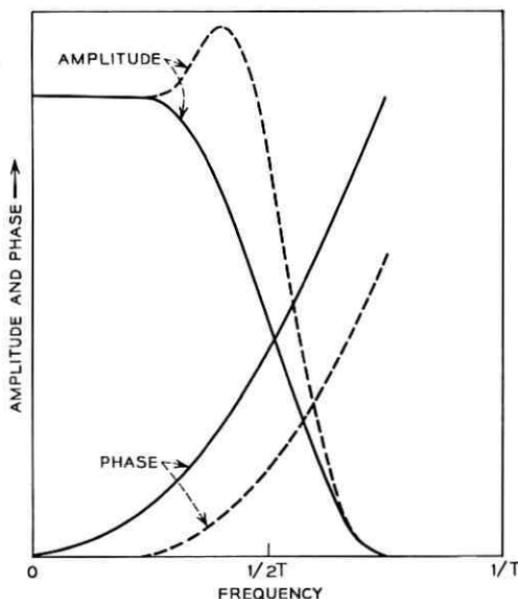


Fig. 16 — Equalization of linear delay distortion.

$$\hat{A}_{x_0}(\omega) = x_0 + \sum_{n=1}^{\infty} (-1)^n (x_n + x_{-n}) \cos n\omega T \quad (100)$$

$$\hat{A}_{y_0}(\omega) = \sum_{n=1}^{\infty} (-1)^n (x_n - x_{-n}) \sin n\omega T. \quad (101)$$

The other components, \hat{A}_{x_0} and \hat{A}_{y_0} , are of no concern to us, since they affect neither $\{x_n\}$ nor the equalizer settings. Now calculate $\hat{C}(\omega)$ from (93) and evaluate its Fourier transform at time nT . Each of these values may be identified with tap settings through (96) and (97). This procedure may seem complex, but it is by far the fastest way to calculate tap settings for long transversal filters on a digital computer. Time-domain techniques involve the solution of simultaneous equations or iterative methods depending on the input pulse. The time required for this type of minimization usually increases with the square of the number of taps involved. As each new tap is added, all the previous tap settings change to readjust for the new minimum for the increased filter length.

On the other hand, using the frequency-domain technique the time required for computing tap settings increases linearly with the number of taps involved and each new tap setting does not change any of the

previous settings. The catch is that the settings calculated by the frequency-domain method apply only to an infinite-length line. For lines of finite length the settings calculated by frequency- and time-domain methods will differ. The time-domain settings are calculated on the basis of minimum distortion — a quantity defined on the basis of time samples. The frequency-domain settings for a finite line are a root-mean-square approximation to the final equalizer frequency characteristic for an infinite line. As the number of taps goes to infinity the tap settings computed by each method approach the same value. For any finite line the time-domain settings are better (lead to less distortion), since they are by definition optimum. However, for equalizers of about 12 taps or more the results are practically the same. (Indeed, one should remember that the distortion criterion itself is somewhat arbitrary.) Thus the frequency-domain technique described here is a fast and accurate way of computing tap settings for long equalizers. Unfortunately the method does not lend itself to implementation, since it involves a number of difficult operations, i.e., division, Fourier inversion.

VI. ACKNOWLEDGMENTS

The author is particularly indebted to F. K. Becker and E. Port, who are largely responsible for the equalization system shown in Fig. 4. The counter-attenuators illustrated in Fig. 5 were designed by L. N. Holzman.

APPENDIX A

Statement and Proof of Theorem I

Theorem I

If $D_0 < 1$, then the minimum distortion D must occur for those N tap gains which simultaneously cause $h_n = 0$ for all $n \in K_N$, $n \neq 0$.

Proof

We prove this theorem by assuming a minimum has occurred at some point other than that specified in the theorem and showing this assumption leads to a contradiction. Thus we assume that D is at a minimum for some sequence of tap gains $\{c_j\}$ and that $h_k \neq 0$ for some $k \in K_N$, $k \neq 0$. Now we will show that there exists another sequence $\{c_j^*\}$ for which $D^* < D$ and hence the contradiction.

Let

$$\begin{aligned} c_j^* &= c_j \quad j \in K_N, j \neq 0, j \neq k \\ c_k^* &= c_k - \Delta \operatorname{sgn} h_k \end{aligned} \quad (102)$$

where

$$\frac{1}{2} |h_k| > \Delta > 0 \quad (103)$$

which is possible, since by hypothesis $h_k \neq 0$. The value of distortion corresponding to the tap gains $\{c_j^*\}$ is, from (14)

$$\begin{aligned} D^* &= \sum_{n=-\infty}^{\infty} \left| \sum_{\substack{j \in K_N \\ j \neq k}} c_j (x_{n-j} - x_n x_{-j}) + x_n \right. \\ &\quad \left. + (c_k - \Delta \operatorname{sgn} h_k) (x_{n-k} - x_n x_{-k}) + x_k \right|. \end{aligned} \quad (104)$$

Using (13) this reduces to

$$D^* = \sum_{n=-\infty}^{\infty} \left| h_n - \Delta \operatorname{sgn} h_k (x_{n-k} - x_n x_{-k}) \right| \quad (105)$$

$$\begin{aligned} D^* &= \sum_{\substack{n=-\infty \\ n \neq k}}^{\infty} \left| h_n - \Delta \operatorname{sgn} h_k (x_{n-k} - x_n x_{-k}) \right| \\ &\quad + \left| h_k - \Delta \operatorname{sgn} h_k (1 - x_k x_{-k}) \right|. \end{aligned} \quad (106)$$

We concentrate for the moment on the second term of (106) and use $h_k = |h_k| \operatorname{sgn} h_k$ to obtain

$$\left| h_k - \Delta \operatorname{sgn} h_k (1 - x_k x_{-k}) \right| = \left| |h_k| - \Delta (1 - x_k x_{-k}) \right|. \quad (107)$$

But $|x_k x_{-k}| < 1$ since

$$D_0 = \sum_{n=-\infty}^{\infty} |x_n| < 1. \quad (108)$$

Therefore $2 > (1 - x_k x_{-k}) > 0$ and, by (103),

$$\left| |h_k| - \Delta (1 - x_k x_{-k}) \right| > 0. \quad (109)$$

Thus we are able to drop the absolute value brackets around the second term in (106) and write

$$\begin{aligned} D^* &= \sum_{\substack{n=-\infty \\ n \neq k}}^{\infty} \left| h_n - \Delta \operatorname{sgn} h_k (x_{n-k} - x_n x_{-k}) \right| \\ &\quad + \left| |h_k| - \Delta (1 - x_k x_{-k}) \right| \end{aligned} \quad (110)$$

$$D^* \leq \sum'_{n=-\infty}^{\infty} |h_n| + \Delta \sum'_{\substack{n=-\infty \\ n \neq k}}^{\infty} |x_{n-k} - x_n x_{-k}| - \Delta(1 - x_k x_{-k}) \quad (111)$$

$$D^* \leq D + \Delta \left\{ \sum'_{\substack{n=-\infty \\ n \neq k}}^{\infty} |x_{n-k}| + |x_{-k}| \sum'_{\substack{n=-\infty \\ n \neq k}}^{\infty} |x_n| - (1 - x_k x_{-k}) \right\} \quad (112)$$

$$D^* \leq D + \Delta \{ D_0 - |x_{-k}| + |x_{-k}| (D_0 - |x_k|) - 1 + |x_k| |x_{-k}| \} \quad (113)$$

$$D^* \leq D - \Delta(1 - D_0)(1 + |x_{-k}|) \quad (114)$$

and since $D_0 < 1$

$$D^* < D \quad (115)$$

which is a contradiction, since we assumed that D was originally at a minimum. This completes the proof of Theorem I.

APPENDIX B

Derivation of Equalizer Frequency Response in Terms of Unequalized Channel Frequency Response

The equalizer frequency response taken about the Nyquist frequency π/T is

$$\hat{C}(\omega) = \hat{C}_{xe}(\omega) + j\hat{C}_{yo}(\omega) \quad (116)$$

where we use the subscripts "x" and "y" to denote real and imaginary components and the subscripts "e" and "o" to denote even and odd components. Breaking the unequalized channel frequency response similarly into its components gives

$$\hat{A}(\omega) = \hat{A}_{xe}(\omega) + \hat{A}_{xo}(\omega) + j\hat{A}_{ye}(\omega) + j\hat{A}_{yo}(\omega) \quad (117)$$

with

$$\hat{A}_{xe}(\omega) = \frac{1}{2}[\hat{A}_x(\omega) + \hat{A}_x(-\omega)] \quad (118)$$

$$\hat{A}_{xo}(\omega) = \frac{1}{2}[\hat{A}_x(\omega) - \hat{A}_x(-\omega)] \quad (119)$$

etc.

We are now ready to put together the equalizer response and the

channel response to get the equalized channel response. Using (117) and (116) for this purpose we obtain

$$\begin{aligned} \hat{A}(\omega)\hat{C}(\omega) &= [\hat{A}_{xe}(\omega) + \hat{A}_{xo}(\omega)]\hat{C}_{xe}(\omega) - [\hat{A}_{ye}(\omega) \\ &+ \hat{A}_{yo}(\omega)]\hat{C}_{yo}(\omega) + j\{\hat{A}_{xe}(\omega) \\ &+ \hat{A}_{xo}(\omega)\}\hat{C}_{yo}(\omega) + \{\hat{A}_{ye}(\omega) + \hat{A}_{yo}(\omega)\}\hat{C}_{xe}(\omega)\}. \end{aligned} \quad (120)$$

This equation must meet the conditions (82) and (83). If δ_o is an arbitrary odd function, condition (82) dictates that

$$\text{Re} [\hat{A}(\omega)\hat{C}(\omega)] = (T/2) + \delta_o \quad (121)$$

$$\begin{aligned} [\hat{A}_{xe}(\omega) + \hat{A}_{xo}(\omega)]\hat{C}_{xe}(\omega) - [\hat{A}_{ye}(\omega) + \hat{A}_{yo}(\omega)]\hat{C}_{yo}(\omega) \\ = (T/2) + \delta_o \end{aligned} \quad (122)$$

$$\hat{A}_{xe}(\omega)\hat{C}_{xe}(\omega) - \hat{A}_{yo}(\omega)\hat{C}_{yo}(\omega) = T/2 \quad (123)$$

$$\hat{A}_{xo}(\omega)\hat{C}_{xe}(\omega) - \hat{A}_{ye}(\omega)\hat{C}_{yo}(\omega) = \delta_o. \quad (124)$$

Since δ_o is an arbitrary function, (124) is automatically satisfied.

Let δ_e be an arbitrary even function. Condition (83) now requires

$$\text{Im} [\hat{A}(\omega)\hat{C}(\omega)] = \delta_e \quad (125)$$

$$[\hat{A}_{xe}(\omega) + \hat{A}_{xo}(\omega)]\hat{C}_{yo}(\omega) + [\hat{A}_{ye}(\omega) + \hat{A}_{yo}(\omega)]\hat{C}_{xe}(\omega) = \delta_e. \quad (126)$$

The even and odd equalities from (126) are respectively

$$\hat{A}_{xo}(\omega)\hat{C}_{yo}(\omega) + \hat{A}_{ye}(\omega)\hat{C}_{xe}(\omega) = \delta_e \quad (127)$$

$$\hat{A}_{xe}(\omega)\hat{C}_{yo}(\omega) + \hat{A}_{yo}(\omega)\hat{C}_{xe}(\omega) = 0. \quad (128)$$

Again (127) is satisfied trivially and we are left with two equations, (128) and (123), in the two unknowns $\hat{C}_{xe}(\omega)$ and $\hat{C}_{yo}(\omega)$. These equations may be solved simultaneously and we arrive at

$$\hat{C}_x(\omega) = \frac{\frac{T}{2} \hat{A}_{xe}(\omega)}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)} \quad (129)$$

$$\hat{C}_y(\omega) = \frac{-\frac{T}{2} \hat{A}_{yo}(\omega)}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)}. \quad (130)$$

REFERENCES

1. Blumlein, A. D., British Patent No. 517516.
2. Kallman, Heinze E., Transversal Filters, Proc. IRE, 28, July, 1940.
3. Kac, M., Random Walk and the Theory of Brownian Motion, in *Selected Papers on Noise and Stochastic Processes*, Nelson Wax, ed., Dover Press, New York, N. Y.
4. Nyquist, H., Certain Topics in Telegraph Transmission Theory, Trans. AIEE, 47, 1928.

Index Reduction of FM Waves by Feedback and Power-Law Nonlinearities

By V. E. BENEŠ

(Manuscript received April 17, 1964)

Feedback systems which achieve subharmonic response by use of power-law nonlinearities and trigonometric identities are described, and the problem of "modes," i.e., multiple responses to the same excitation, in some of these systems is discussed. It is shown, by a thorough discussion of an example, that suitable choices of a trigonometric identity and a loop filter can be made which lead to locally asymptotically stable subharmonic orbits. An application to FM demodulation is suggested: the subharmonic modes can be used to reduce the index of an input wide-index wave so that a narrower IF filter than for conventional FM suffices, as in the FM demodulator with feedback, but without a controlled oscillator or even a mixer.

I. INTRODUCTION

As is known, it is possible to design feedback systems that reduce the index of an FM wave by an explicit use of power-law nonlinearities based on certain simple trigonometric identities. Circuits which use feedback to achieve subharmonic operation, and which are stable under changes in the input frequency over at least a limited range, have been built, tested, and described in the literature.¹ They have the behavior predicted by the trigonometric identities. Some of them depend on nonlinear conversion of a signal containing harmonics of θ into one containing only the first harmonic, and others depend on the inverse process of generating harmonics.

Our purpose is to discuss the problem of "modes" (i.e., different responses to the same excitation) in these systems, and the problem of the stability of the interesting subharmonic modes. We also indicate an application to frequency modulation: by incorporating the principle on which the systems are based into a frequency modulation receiver, one obtains circuits with some of the properties and advantages of the FM demodulator with feedback^{2,3} proposed by J. G. Chaffee. However,

none of the circuits suggested here contains a voltage-controlled oscillator, and some do not even contain a mixer.

To illustrate the principles involved, let us consider the trigonometric identity¹

$$\sin 3\theta = 3 \sin \theta - 4 \sin^3 \theta.$$

If we set

$$\psi(x) = \text{sgn } x \cdot |x|^{1/3}$$

the above identity can equally well be put in the form

$$\psi(3 \sin \theta - \sin 3\theta) = 4^{1/3} \sin \theta.$$

Thus if we view, as in Fig. 1, $4^{1/3} \sin \theta$ as the output of the nonlinearity $\psi(\cdot)$, and feed this back through a gain of $3 \cdot 4^{-1/3}$ to an adder whose input is

$$-\sin 3\theta,$$

we will have a feedback system which is driven by $-\sin 3\theta$, and which produces $\sin \theta$, a 3-to-1 reduction of index of modulation if the angle 3θ is taken to be of the form

$$3\theta(t) = \omega t + 3\varphi(t), \quad \varphi = \text{signal}.$$

It is apparent¹ that other trigonometric identities can be used in analogous fashion to get an n -to-1 index reduction, with n any positive integer ≥ 2 . For example, with

$$\psi(x) = \text{sgn } x \cdot |x|^{1/5}$$

we have

$$\psi(\cos 5\theta + 20 \cos^3 \theta - 5 \cos \theta) = 16^{1/5} \cos \theta,$$

corresponding to the system of Fig. 2.

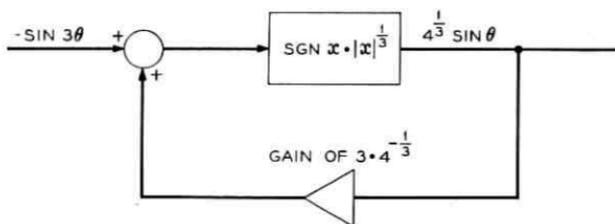


Fig. 1 — System using 1/3-power nonlinearity.

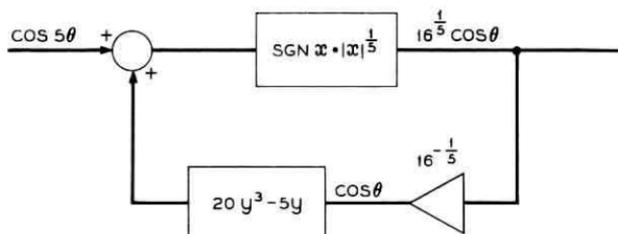


Fig. 2 — Loop based on $1/5$ -power nonlinearity.

Another system, based on the identity

$$4 \cos^3 \theta = \cos 3\theta + 3 \cos \theta,$$

is shown in Fig. 3. It generates the “compressed” signal $-3 \cos \theta$ by subtracting $4 \cos^3 \theta$ from the wide-index signal $\cos 3\theta$.

II. FILTERING

If the principles illustrated above are to be used in a communications receiver, it is probably desirable to perform some filtering to remove undesired components of noise or feedback signal. Thus in practice the feedback loop would (e.g.) include a filter which removed all components not in the (essential) band of $\sin \theta(\cdot)$. For many filters, and choices of input phase $\theta(\cdot)$, presence of the filter will of course mean that the signal in the loop is no longer so simply related to $\theta(\cdot)$ as it was in the examples above: use of the trigonometric identity to relate the loop signal to the input may be inexact. However, there exist filters and choices of $\theta(\cdot)$ for which this does not occur. (See Ref. 5, and Section V herein.) In any case, if the filter passes $\sin \theta(\cdot)$ without essential distortion the identity will remain true for practical purposes. For example, with

$$\psi(x) = \operatorname{sgn} x \cdot |x|^{1/3}$$

again, and the identity

$$\psi(\cos 3\theta + 3 \cos \theta) = 4^{1/3} \cos \theta$$

we would follow $\psi(\cdot)$ with a filter that passed $\cos \theta$ but removed out-of-band noise, and get a system like that of Fig. 4.

III. THE POSSIBILITY OF SEVERAL “MODES”

It has been pointed out in the literature⁴ that certain frequency di-

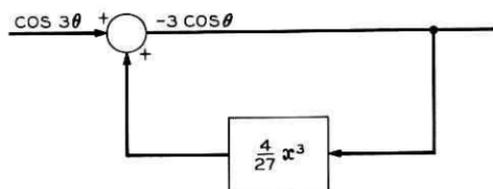


Fig. 3 — System based on cubic nonlinearity in the feedback.

viders based on regeneration and modulation are not necessarily “self-starting,” and that sizeable starting voltages may be needed to set them off. Put in the language of differential equations, this means for example that there may be two entirely different kinds of steady-state response (to the same steady-state signal), one (say) oscillating with large amplitudes, and the other taking place in a region of asymptotic stability around a critical point, with small oscillations. The “starting voltage” is needed to get the system out of the region of stability.

A. J. Giger has suggested that a similar situation will obtain in realizations of the circuits described above, even though they contain an adder rather than a multiplier (modulator). Clearly, whether a realization is self-starting is going to depend on the circuit details as well as on the principles at issue here, and each case will have to be studied on its own merits. However, these general remarks are pertinent:

(i) The non-self-starting frequency divider is just a special case of the well-known but incompletely understood phenomenon that a system may not have a unique asymptotic response. It is a property specific to the systems described so far that they depend on a fractional-power nonlinearity, and as a result it is possible, though not necessary, that they fail to have unique periodic responses to some periodic signals.⁵

(ii) The desired operation of the systems proposed above depends on evoking a suitable *subharmonic* response. It is known⁵ that not all solutions need contain such components of lower frequency than the input.

(iii) It is also known⁵ that even when a subharmonic periodic solution exists, it is itself only unique up to certain phase shifts. Specifically, if we obtain a solution with a component $e^{\pi i \omega t}$ when the input contains only harmonics of $e^{\pi i n \omega t}$, $n > 2$, then any *translation* of a solution by $2k/n$ for $k = 1, \dots, n$ is also a solution.

For example, the steady-state response of the circuit of Fig. 4 might or might not contain the desired subharmonics, and thus it might not work as planned unless care is taken to ensure that it slips into the right “mode” of operation initially, and that this mode is stable under the

perturbations due to the presence of signal and noise. The mathematical analysis of such phenomena is often arduous and, in many cases, quite incomplete. From a practical point of view the mode problem (when it occurs) might best be resolved by encouraging the desired mode by pulsing the (tuned circuit) filter, by adding automatic gain control features that cut in when the amplitudes of the desired mode are low, or by designing the filter to have zeros of transmission at certain values of frequency associated with the undesired modes, such as their fundamentals.

However, while it is necessary to emphasize that the problems mentioned above exist, it is also important to state that the picture is not all black: known methods of analysis and design suffice to ensure local asymptotic stability of some of the subharmonic modes described above. An example is worked out in some detail in Section V.

IV. BIAS

For theoretical reasons it may be undesirable, and for practical reasons impossible, to use a nonlinear characteristic which has an infinite slope at the origin. For example the singular nature of the fractional-power nonlinearities at the origin may preclude a conventional closed-loop, open-loop stability analysis by linearization around that point. Or, if the nonlinearities are being obtained by the use of diodes, such a slope is physically unattainable. These difficulties, along with most passages of the system through or near such a singular point, can be avoided by the addition of what amounts in electrical terms to "dc bias" at various points of the system, in such a way as to (roughly) move the operating point of the system to a desired region of the nonlinear characteristic $\psi(\cdot)$. In this region $\psi(\cdot)$ might be of Lipschitz character, or it might be particularly well represented by a particular diode. Such biasing can also be used to eliminate some equilibrium points of the system, and thus to reduce the number of solutions; it can also be used to increase

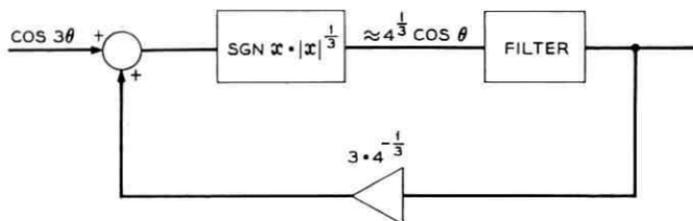


Fig. 4 — Filtering in a loop using 1/3-power nonlinearity.

the size of certain regions of asymptotic stability, and thereby enhance the dynamic stability of desired solutions. It therefore furnishes substantial latitude for design.

Thus, e.g., to use the identity

$$\sin^2 \theta = \frac{1 - \cos 2\theta}{2},$$

we can multiply by $b^2 \neq 0$ and add

$$a^2 + 2ab \sin \theta$$

to both sides, so that

$$(a + b \sin \theta)^2 = a^2 + (b^2/2) + 2ab \sin \theta - (b^2/2) \cos 2\theta.$$

Choosing $|a| > |b|$ ensures that the right-hand side is bounded away from zero. Taking the square root of both sides, we base design on the "biased" identity

$$a + b \sin \theta = (a^2 + \frac{1}{2}b^2 + 2ab \sin \theta - \frac{1}{2}b^2 \cos 2\theta)^{1/2}. \quad (1)$$

Discussion of an example of design based on (1) follows.

V. EXAMPLE, WITH ANALYSIS OF ORBITAL ASYMPTOTIC STABILITY

We now consider the circuit depicted in Fig. 5, with

$$\psi(x) = \operatorname{sgn} x \cdot |x|^{1/2},$$

with input

$$y(t) = a^2 + \frac{1}{2}b^2 - \frac{1}{2}b^2 \cos 2t,$$

and a filter whose impulse response $k(\cdot)$ is integrable and has a Fourier transform

$$K(\omega) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} k(t)e^{-i\omega t} dt$$

such that $K(0) = 0$ and $K(1) = (2\pi)^{-1/2}(2a)$. For our example we use the second-order filter

$$K(\omega) = \frac{(2\pi)^{-1/2} 2ai\omega}{c(i\omega)^2 + i\omega + c}, \quad (2)$$

with $c > 0$, so that poles of $K(\cdot)$ are in the left half-plane.

It is now easy to show, using the theory of Fourier series, that the system of Fig. 5 has a subharmonic response

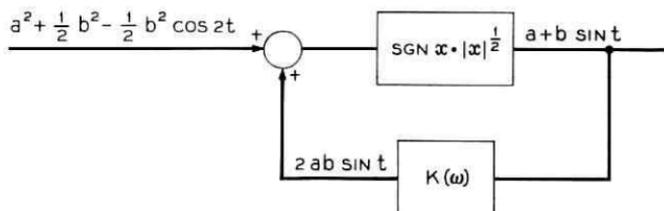


Fig. 5 — Example for orbital asymptotic stability.

$$x(t) = a^2 + \frac{1}{2}b^2 + 2ab \sin t - \frac{1}{2}b^2 \cos 2t, \quad (3)$$

where $x(\cdot)$ is the input to the nonlinearity. The loop equation for this solution (with no transients!) is simply

$$x(t) = y(t) + \int_{-\infty}^{\infty} k(t-u) \operatorname{sgn} x(u) |x(u)|^{1/2} du.$$

Substituting the expression that (3) gives for $x(\cdot)$, using the identity (1), shows that the filter removes the dc term a from the output of the nonlinearity and multiplies the amplitudes of frequencies ± 1 by $2a$. We remark that the constant $c > 0$ in (2) does not occur in the subharmonic solution (3), and that the constant b does not occur in the transfer function $K(\omega)$. Also, for the particular input we chose, the presence of the filter does not render the trigonometric identity being used inexact.

As is well known, for fixed values of a and c there are many ways of describing the circuit of Fig. 5 by differential equations so as to give rise to (2) as the transfer function of the filter. For some of these ways the solution $x(\cdot)$ found above may be orbitally asymptotically stable, for others it may not. In short, the stability of this subharmonic solution probably depends on the way in which the action of the transfer function is represented by differential equations. We shall consider representations of the form

$$\dot{z} = Az + \delta\psi(\beta'z + y(t)), \quad (4)$$

where $z(\cdot)$ is a 2-vector valued function, β is a 2-vector, δ is a 2-vector proportional to the unit 2-vector, and A is a stable 2×2 matrix. In this representation the periodic solution $x(\cdot)$ of (3) will have the form

$$x(t) = \beta'z(t) + y(t).$$

These representations and their simple properties are used merely as an illustration, because they readily admit an analysis of the asymptotic

stability of the orbit corresponding to the solution $x(\cdot)$. They do not come near exhausting the possibilities for finding stable subharmonic forced oscillations in feedback systems.

Since for $t_2 \geq t_1$, (4) gives

$$\beta'z(t_2) = \beta'[\exp A(t_2 - t_1)]z(t_1) + \int_{t_1}^{t_2} \beta'[\exp A(t_2 - u)]\delta\psi(\beta'z(u) + y(u))du$$

it is clear that (4) will represent the circuit if and only if

$$\beta'(i\omega I - A)^{-1}\delta = (2\pi)^{1/2}K(\omega) = \frac{2ai\omega}{c(i\omega)^2 + i\omega + c},$$

i.e., if and only if both

$$\det(i\omega I - A) = (i\omega)^2 + c^{-1}i\omega + 1, \\ \beta_1 \begin{vmatrix} \delta_1 & -a_{12} \\ \delta_2 & i\omega - a_{22} \end{vmatrix} + \beta_2 \begin{vmatrix} i\omega - a_{11} & \delta_1 \\ -a_{21} & \delta_2 \end{vmatrix} = 2ai\omega/c, \quad (5)$$

The first condition is met if

$$a_{11}a_{22} - a_{12}a_{21} (= \det A) = 1, \\ a_{11} + a_{22} = -c^{-1}.$$

The second condition is equivalent to $\beta'\delta = 2ac^{-1}$ and $\beta'A^{-1}\delta = 0$, taken together. Since, to facilitate stability analysis for this example, we wish to impose $\delta_1 = \delta_2 \neq 0$, the latter condition is $\beta'A^{-1}\mathbf{1} = 0$, $\mathbf{1} =$ unit 2-vector. All these conditions together can be met in many ways. A convenient choice is $\beta = (2a/c, 0)'$, $\delta = 1$, $a_{22} = -a_{21}$, $(a_{11} - a_{12})a_{22} = 1$, and

$$a_{11}(-a_{11} - c^{-1}) - a_{12}(-a_{11} - c^{-1}) = 1,$$

the last imposing a rational relation between a_{11} and a_{12} , thus leaving one parameter still free.

It can be verified that when the conditions (5) above are met, then $\beta'z(t) = 2ab \sin t$ does define a periodic subharmonic orbit of the differential system (4). We are now in a position to make a linear local asymptotic stability analysis for the subharmonic orbit $\beta'z(t) = 2ab \sin t$, by Lyapunov's classical theorem.⁶ The (periodic) linearization matrix is

$$A + \text{diag } \delta f(t)$$

where

$$\begin{aligned}
 f(t) &= \psi'(a^2 + 2ab \sin t + \frac{1}{2}b^2 - \frac{1}{2}b^2 \cos 2t) \\
 &= \psi'(x(t)) \\
 &= \frac{1}{2\psi(x(t))} \\
 &= \frac{1}{2}(a + b \sin t)^{-1}.
 \end{aligned}$$

If $\text{diag } \delta = dI$, d a scalar, it is easily verified that the fundamental matrix associated with this periodic matrix is

$$\Phi(t) = e^{At}u(t)$$

where $u(\cdot) > 0$ is defined by

$$\begin{aligned}
 u_0 &\equiv 1 \\
 u_{n+1}(t) &= d \int_0^t f(v)u_n(v)dv \\
 u(t) &= \sum_{n=0}^{\infty} u_n(t).
 \end{aligned}$$

We are therefore interested in the characteristic values of $\Phi(2\pi)$, i.e.,

$$e^{2\pi A}u(2\pi).$$

These are of the form

$$u(2\pi)e^{2\pi\mu}$$

where μ is a characteristic value of A , and so we may conclude at once⁶ that if

$$-\text{Re } \mu = \frac{1}{2c} > \frac{1}{2\pi} \log u(2\pi),$$

then the orbit determined by $\beta'_z(t) = 2ab \sin t$ is locally asymptotically stable, i.e., there exists a neighborhood of it from which all solutions approach the orbit.

Clearly

$$u(t) \leq \exp \left\{ t \sup_{0 \leq u \leq t} |d| \cdot |f(u)| \right\}$$

whence

$$\frac{1}{2\pi} \log u(2\pi) \leq \frac{1}{2} \frac{|d|}{a - |b|}$$

so that the inequality

$$a > c |d| + |b|$$

suffices for local asymptotic stability of the orbit.

Note that changing the sign of b corresponds to changing the sign in all the odd components in the periodic solution $x(t)$ found above; the resulting function is also a periodic solution differing from $x(t)$ only in phase, by exactly π .

VI. APPLICATION TO AN FM DEMODULATOR WITH FEEDBACK

The examples of index reduction described in the preceding sections suggest that it is possible to design FM demodulators with frequency feedback that contain no voltage-controlled oscillators and even no mixers. Several methods of realizing such a possibility will be discussed, based on the principles exemplified.

The simplest demodulation scheme of this sort is obtained by a specialization of the system of Fig. 4. We merely specify that the filter have a narrow passband centered around an intermediate frequency ω that is $\frac{1}{3}$ the carrier frequency, and that it introduce negligible amplitude variations for signals in the passband, so as not to interfere unduly with the trigonometric identity. This gives rise to the demodulator of Fig. 6, in which $\theta(\cdot)$ has the form

$$\theta(t) = \omega t + \varphi(t),$$

with φ a baseband signal. The feedback, made at IF, reduces the modulation index 3-to-1. Here the carrier frequency is *three* times the intermediate frequency, but this relationship can easily be changed by remodulating or, for that matter, by using a different trigonometric identity as the basis of design. We note that no mixer, and no voltage-controlled oscillator, is used. Also, the phase of the signal fed back is crucial: excessive phase shift in the filter is as intolerable here as in conventional FM with feedback.

As a demodulator, the circuit of Fig. 6 shares with Chaffee's circuit the advantage that the wide band of noise which must be passed by the initial amplifier along with the wide-index signal is not admitted to the detector. This circumstance is important, because a principal object of feedback (in FM with feedback) is to reduce the noise level at the detector by filtering all but a small part of the noise. However, it remains to be seen how well the nonlinearity $\text{sgn } x \cdot |x|^{1/3}$ performs its function in the presence of the wideband noise that enters it, since the resulting amplitude modulation $a(\cdot)$ at the input renders the trigonometric identity being used here inexact. This AM due to noise might be removed

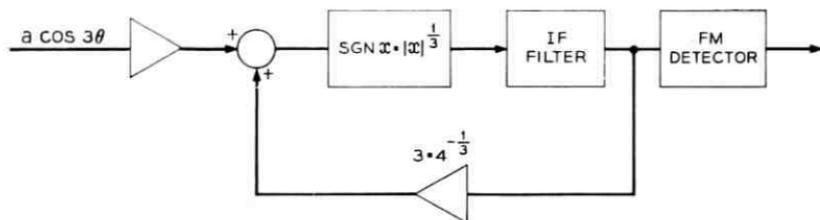


Fig. 6 — Possible utilization of $1/3$ -power nonlinearity in FM demodulator.

by inserting a limiter and filter to follow the IF filter, as in conventional systems, at the expense of incurring additional phase shift.

A final point, due to A. J. Giger, is that, unlike Chaffee's circuit, the present one *retains* the carrier phase instead of discarding it and operating independently of it.

VII. CIRCUITS WITHOUT FRACTIONAL-POWER NONLINEARITIES

It is straightforward to generate other, quite different designs based on the same identity,

$$\cos 3\theta = 4 \cos^3 \theta - 3 \cos \theta,$$

designs which do not depend on a *fractional*-power nonlinearity, and so do not incur the problems above. For example, Fig. 7 shows a design very much like that of the conventional FM with feedback demodulator, except that the detector-controlled oscillator is replaced by the nonlinearity

$$4y^3 - 3y.$$

Applied to $\cos \theta$, this gives $\cos 3\theta$, to be used as the feedback input to the mixer. If the other mixer input, i.e., the incoming signal, is the

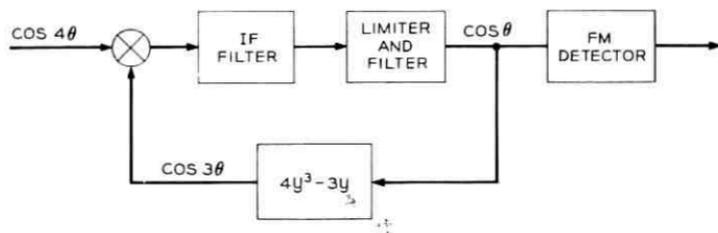


Fig. 7 — Cubic nonlinearity used in FM feedback system with modulator.

wide-index signal $\cos 4\theta$, the filter can be made to select the difference frequency component $\cos \theta$ to complete the loop and provide the feedback signal. In this system the carrier frequency is four times the intermediate frequency.

A particularly simple circuit, based on Fig. 3 and using only an adder, is depicted by Fig. 8. In this design the feedback is through the simple cubic nonlinearity

$$\frac{4}{27} x^3.$$

If $-3 \cos \theta$ is applied to this, and the output is combined in the adder with an incoming wide-index signal $\cos 3\theta$, the adder output is $-3 \cos \theta$. This passes substantially unchanged through the narrow IF filter suitable for the low-index wave $\cos \theta$, while of the wide band of noise accompanying $\cos 3\theta$ at the input only a narrow band can pass the IF filter. For practical purposes, the cubic characteristic would only be required over the range $|x| \leq 3$, and standard stability analyses can be used. Again, carrier phase is retained.

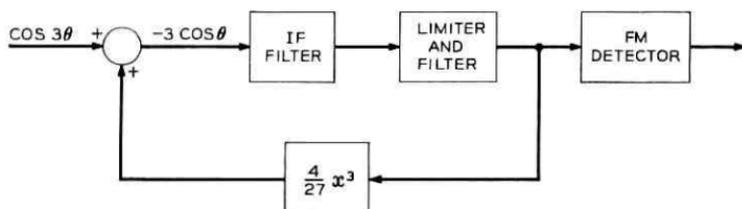


Fig. 8 — Cubic nonlinearity used in FM feedback system with adder.

VIII. ACKNOWLEDGMENTS

The author is pleased to thank S. Darlington, A. J. Giger, and I. W. Sandberg for extensive, helpful discussion.

REFERENCES

1. Hughes, W. L., *Nonlinear Electrical Networks*, Ronald Press, New York, 1960.
2. Chaffee, J. G., The Application of Negative Feedback to Frequency Modulation Systems, Proc. IRE, 27, 1939, pp. 317-331.
3. Giger, A. J., and Chaffee, J. G., The FM Demodulator with Negative Feedback, B.S.T.J., 42, 1963, pp. 1109-1135.

4. Hales, E. B., and Holdam, J. V., Sinusoidal Frequency Multipliers and Dividers, in *Waveforms*, Vol. 19 of M.I.T. Radiation Laboratory Series, McGraw-Hill, New York, 1949, p. 19.
5. Beneš, V. E., and Sandberg, I. W., Applications of a Theorem of Dubrovskii to the Periodic Responses of Nonlinear Systems, *B.S.T.J.*, *43*, 1964, pp. 2855-2872.
6. Pontryagin, L. S., *Ordinary Differential Equations*, Addison-Wesley, Reading, Mass., 1962, p. 264, Theorem 25.

Generation of Orbital Elements for the TELSTAR[®] Communications Satellites

By L. C. THOMAS

(Manuscript received December 8, 1964)

The technique now regularly in use for the generation of orbital elements for the TELSTAR communications satellites is described using angle-only and/or angle-range data versus time as input information. It is found that secular perturbation considerations are sufficient to permit trajectory prediction with pointing errors of about 0.05° over 100 orbits from a single set of elements. Modified orbital elements are chosen as the orbit description, since they explicitly express the secular rates and thereby simplify and reduce the cost of drive tape generations for the Andover ground station. The rates are derived both from perturbation theory and from direct measurement. The size and shape accuracy of the predicted orbit ellipse is improved by statistical means using trajectory data from a number of passes over a particular ground station. This permits better round-the-world predictions. Finally the computer-operator ensemble presently used for generating the orbits of the TELSTAR satellites is described.

TABLE OF CONTENTS

I. INTRODUCTION.....	604
II. EFFECT OF DATA ON ANALYSIS.....	605
III. THE MODIFIED ORBITAL ELEMENTS.....	607
IV. TRAJECTORY PREDICTIONS.....	608
V. DETERMINING SATELLITE RANGE FROM ANGLE DATA.....	610
5.1 <i>A Brief History</i>	610
5.2 <i>General Procedure</i>	612
5.3 <i>Synthesizing the Range</i>	612
VI. GENERATING THE ORBITAL ELEMENTS.....	621
6.1 <i>General Procedure</i>	621
6.2 <i>Generating the Unperturbed Elements</i>	622
6.2.1 <i>Locating the Subsatellite Point</i>	622
6.2.2 <i>Computing Orbit Inclination and Ascending Node</i>	623
6.2.3 <i>Computing the True Anomaly Along with Orbit Size and Shape</i>	625
6.2.4 <i>Reduction to Epoch</i>	628
6.2.5 <i>Anomalistic Period</i>	629
6.3 <i>Adding the Perturbations to the Orbital Elements</i>	630
VII. THE SECULAR PERTURBATIONS.....	630
7.1 <i>Perturbation Theory and Formulae</i>	630
7.2 <i>Determining the Anomalistic Period from the Nodal Period</i>	632

VIII. ROUND-THE-WORLD ORBIT PREDICTIONS.....	634
IX. IMPROVING THE ELEMENT RATES BY DIRECT MEASURE.....	636
9.1 <i>Nodal Regression</i>	636
9.2 <i>Anomalistic Period</i>	637
9.3 <i>Apsidal Advance Rate</i>	638
X. A COMPUTER-OPERATOR ENSEMBLE FOR ORBIT GENERATION.....	639
10.1 <i>The Computer Programs</i>	639
10.1.1 <i>The MOEGEN G3 Program</i>	639
10.1.2 <i>The MOERATE Program</i>	640
10.1.3 <i>The REFINE PERIOD Program</i>	640
10.1.4 <i>The CORD DOPPLER G2 Program</i>	641
10.1.5 <i>The MONITOR TAPE ANALYSIS Program</i>	641
10.2 <i>Orbital Elemental Generation Operational Procedures</i>	641
10.2.1 <i>Initial Element Generation</i>	641
10.2.2 <i>Subsequent Element Generation</i>	645
XI. TYPICAL EXPERIMENTAL RESULTS.....	646
XII. CONCLUSIONS.....	665
XIII. ACKNOWLEDGMENTS.....	666

I. INTRODUCTION

The calculation of elements for orbits similar to those of recent near-earth satellites such as the first and second TELSTAR communication satellites, Echo I and II, the Tiros series and the Syncom satellites is accomplished in this analysis by considering the secular effects of the earth's gravitational potential upon the satellite's orbit. Here, the earth is taken as an oblate spheroid whose field is independent of longitude and is symmetrical about the equatorial plane. In the analysis, the gravitational effects of all bodies other than the earth are ignored, as are atmospheric drag effects and solar light pressure.¹ For orbits at distances in the thousand-mile range, luni-solar attraction can change the satellite height by no more than a few hundred feet. This can alter ground station pointing angles by less than 0.01° , which is below the resolution and tracking accuracy of the horn antenna* at Andover, Maine which supplies the bulk of the basic pointing data for Bell Telephone Laboratories use. The principal planetary perturbation is from Venus, which peaks at 3 to 4 orders of magnitude less than the luni-solar effects. Air drag effects at 1,000 miles are at least an order of magnitude below the solar gravity perturbations.

In the technique to be described, modified orbital elements (osculating elements plus explicit secular perturbations) are derived essentially from three sets of observations of the satellite. If only angle information is available, a modified form of Gauss' method is used to ascertain the range. † If some range information is available, this is added to the analy-

* This antenna has a tracking jitter of about 0.005° . Its beam can resolve the position of a satellite to within about 0.02° .

† It is recognized that the range to an orbiting celestial body cannot be ascertained if the three measured sight lines from observer to the body are coplanar.

sis. If range data for all three observations are on hand, the modified Gauss method is bypassed and elements are calculated from geometry and perturbation theory.

The degree of goodness of elements so obtained is ascertained by comparing resulting predictions of the satellite trajectory with actual data from the same pass as well as with data from passes prior to or after the one being used to generate the elements. It has been this author's experience that the fit of the predicted trajectory to the orbit from which it came is generally within 0 to 0.02° , but that the anomalistic period determination from single-pass data is not sufficiently accurate to hold the errors under 0.1° over more than a few orbits. However, this difficulty is solved by calculating two or more sets of elements from passes separated by several days and combining results to ascertain the period and other secular perturbations more exactly. Table I shows the resulting improvements under such a mode of operation for the second TELSTAR satellite.

It is to be noted that such neglected perturbations as those from the higher-order terms in the earth's potential function or solar radiation pressure may be operationally introduced by not only calculating the period from two passes separated in time, but also computing changes in the other orbital element values. This has been done in the case of Echo II to enable predictions to within 0.1° * over a time of about ten days.

The first portion of this paper presents techniques for developing orbital elements suitable for predicting future satellite pointing angles for the ground station collecting the original data. Later portions of this paper will deal with techniques aimed at producing a good round-the-world orbit fit and element rate improvement. Finally, a computer-operator ensemble used to generate modified orbital elements for the TELSTAR satellites will be described.

II. EFFECT OF DATA ON ANALYSIS

While the use of modern computers makes it entirely possible to record and process hundreds of pointing angles for each pass of a satellite,

In this case such a situation cannot long prevail because of the rapid motions of near-earth satellites and their orbits, as well as the rotation of the ground station on the spinning earth. Even when the orbital inclination equals the geocentric latitude of the station, only for portions of certain passes is the station in the orbit plane. The three-point orbit analysis presented here is therefore not greatly restrictive.

* This is the resolution level of the equipment which recorded the Echo II pointing angles.

TABLE I — IMPROVEMENT OBTAINED BY USING SEVERAL SETS OF ELEMENTS

Period Determined by	Initial Pass	Pointing Errors (Angular Offset)		
		10 Orbits Later	100 Orbits Later	1000 Orbits Later
Data from 1 pass	0.02	0.5°	5°	over 10°
Data from 2 passes separated by 1 day	0.02	0.03°	0.3°	1°
Data separated by 1 week	0.02	0.02	0.05°	0.3°

it is the purpose of this paper to describe techniques which can operate with a sparsity of data. The reduction of the quantity of data to be handled, even by automatic means, is of increasing importance both from a time and cost basis as more and more communications satellites and satellite systems become available. Minimizing the quantity of data is of paramount importance in implementation of small computers at tracking sites which often have very limited storage capacities, long access times, and slow machine cycles. Thus we shall be concerned with generating modified orbital elements from passes having as few as 6 or 12 data points available. Under these conditions, no statistical processing of data prior to their use in orbit generation is advisable or in many cases even sensibly possible. Rather, the procedure is to first generate elements from sets of three pointing angles (and ranges if available) on a pass, and then compare trajectories produced from these elements to the entire original set of data. By appropriate selection of triads, data points unsuitable for orbit prediction may be easily discovered and discarded, even though it may be difficult to preselect data on a purely statistical basis. For use in long-range predictions, orbital element rates (orbital period and other secular perturbations) are determined by comparisons between elements of different epochs, and these rates are used in place of the theoretical perturbations calculated in generations from a single satellite pass.

The tracking data used to determine the orbits of the TELSTAR satellites come from a single ground station. Any increase in station noise or malfunction of trajectory equipment can therefore directly affect the generated elements. Because of this, it was decided not to depend upon an uninterrupted or near uninterrupted flow of "valid" data from the ground station in any analysis but to maintain flexibility by developing elements which would predict trajectories with reasonable accuracy over a number of weeks. Concentration upon the establishment of a con-

sistent set of secular rates derived from elements generated at various epochs seemed a logical approach to this end. These rates are virtually constant for orbiting particles like TELSTAR which lose energy at an extremely low rate and whose oscillatory departures from the secular rates are not serious.

III. THE MODIFIED ORBITAL ELEMENTS

Trajectory generation is simplified by choosing a set of orbital elements in which the only perturbations are secular ones and these are explicitly expressed. This in essence reduces the trajectory prediction process to one of simple geometry. Specifically, the perturbations considered are those which produce nodal regression, perigee advance, and period modifications. They have no secular effect on the orbit inclination, eccentricity, or radius of perigee. Cyclic perturbations are ignored. A further simplification results by quoting the elements for an epoch that is the time when the satellite arrives at perigee.

No satellite moves precisely in an ellipse about the earth, and yet the elements to follow will imply plane elliptical motion. It may be helpful to consider the satellite to be moving along an ellipse in a plane which itself regresses westward due to the earth's oblateness. Superimposed upon this motion is a rotation of the major axis of the orbit ellipse about the earth's center located at one focus of the ellipse. Under these conditions, the orbital elements merely state the instantaneous position of the ellipse in space at the epoch quoted. The modified orbital elements² (commonly referred to as MOES) to be used in this analysis are:

Epoch — the year, month, day, hour and decimal minute at which the elements are quoted. This epoch corresponds to the time that the satellite arrives at perigee with only secular perturbations considered.

Inclination — tilt of the orbit plane in degrees referenced to the earth's equator and measured counterclockwise from the equator to the orbit plane as seen from a point above the ascending node.

West longitude of the ascending node — the west earth longitude of the ascending node of the orbit at epoch in degrees.

Prime sweep interval (PSI) — the apparent orbit regression expressed as a time. Specifically, it is the time required for a point on the earth's equator to pass under the ascending node on two successive occasions. It includes, therefore, the true regression referenced to an inertial frame, plus the apparent regression of almost one degree per day due to the earth's revolution about the sun. Often the prime sweep interval is expressed as one mean solar day (of 1440 minutes) minus M , where M is simply the number of minutes added to one day to yield the PSI.

Argument of perigee — the angle in degrees measured at the earth's center in the direction of the satellite's travel and lying between a radius vector to the ascending node position and one to perigee.

Perigee motion — the secular motion of perigee in degrees along the orbit per anomalistic period. It is positive if perigee moves in the direction of the satellite's motion.

Anomalistic period — the time in minutes required for the secularly perturbed satellite to move from perigee to the next occurring perigee. Generally, during this time the perigee is also moving.

Change in period — the change in the anomalistic period in minutes per anomalistic period.

Eccentricity — the eccentricity of the orbit ellipse. A constant under the assumed conditions of only secular perturbation and no orbit decay.

Radius of perigee — geocentric distance of the satellite at perigee in statute miles.

IV. TRAJECTORY PREDICTION

Since each TELSTAR satellite was launched, it has been a system requirement to predict the position as seen from Andover, Maine on a rather continuous basis. While the demand for predictions of the first TELSTAR satellite trajectories has decreased because that satellite is no longer active, the second TELSTAR satellite requirements have continued. Predictions currently are made for every pass of the second TELSTAR communications satellite according to the methods described herein. As a result of this type of operation it became imperative to develop a computational technique which could rapidly and economically generate trajectories on magnetic tapes capable of pointing an antenna with an accuracy of about 0.05° . The use of modified orbital elements for this purpose permitted such predictions by purely geometric means and without recourse to any integration procedures, numerical or otherwise. The method to be outlined here has been embodied into a program in regular use on an IBM 7094 computer to produce drive tapes for the TELSTAR satellites. The program execution times for a typical hour-long pass in which pointing angles are given every minute is under 0.002 hour and costs about \$1 to compute. The total over-all machine cost of an hour-long drive tape is about \$2, which includes the use of some auxiliary equipment.

The method of predicting pointing angles from a ground station will now be described, using the modified elements as the orbit description. The general procedure will be outlined without explicitly stating all of

the geometry, which involves mathematics no more sophisticated than spherical trigonometry.³

The elements specify the west longitude position of the ascending node and the argument of perigee (P) along with the secular rates so that, for each instant of time, the position of perigee is established as shown in Fig. 1.

Since the anomalistic period is given, the true anomaly V of the satellite must always be known. This, along with the orbit inclination i , describes the subsatellite latitude and longitude as functions of time. Since the radius of perigee, the eccentricity, and the satellite's true anomaly are available, the distance of the satellite above the earth's center is easily determined.

If the latitude, longitude, and height above mean sea level of the ground station are known, it becomes a matter of geometry to compute the azimuth, elevation, and slant range of the satellite from that station. Fig. 2 illustrates this. The great circle arc between SSP and the ground station can be computed to yield the central angle g ; therefore the slant range may be found by simple triangulation, since r and R are known (see Appendix B for R calculations).

A plane tangent to a spherical earth is now inserted at the ground station. Azimuth in this plane is determined by solving the spherical

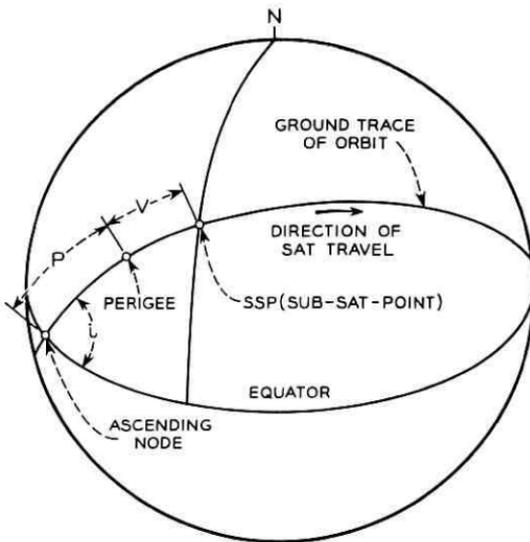


Fig. 1 — Geometry for determining the subsatellite point.

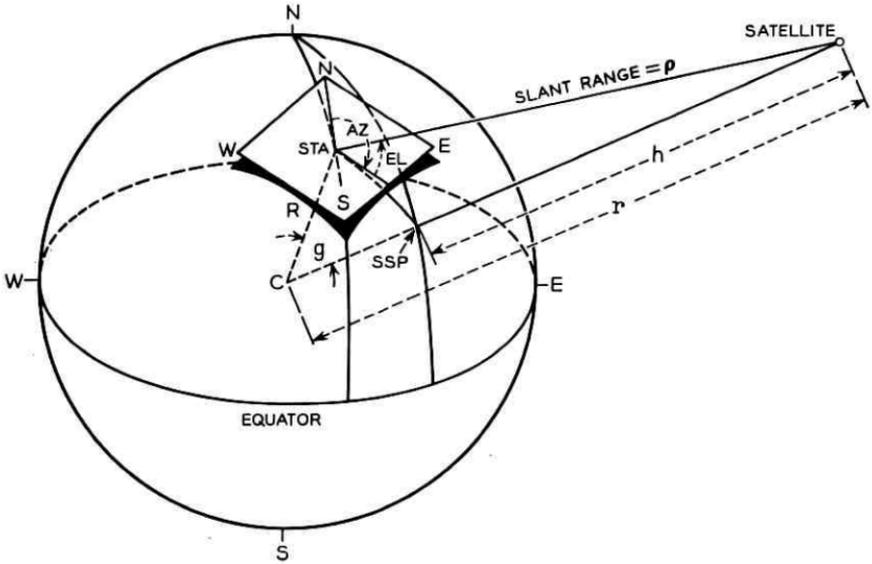


Fig. 2—Satellite azimuth-elevation details.

triangle N , SSP , STA (station). Elevation is computed from the tangent plane perpendicular to R . Predicted azimuth and elevation referred to the oblate earth are derived from the azimuth and elevation related to a spherical earth by tipping the tangent plane north by an appropriate amount depending upon the latitude of the ground station.*

V. DETERMINING SATELLITE RANGE FROM ANGLE DATA

5.1 *A Brief History*⁴

The mathematical determination of orbits for celestial bodies began with Newton in 1680, when he calculated a parabolic orbit for a comet that appeared in that year. This was perhaps made feasible by Doerfel, who suggested that the sun lies at the focus of a cometary orbit. Euler in 1744 added to Newton's work the technique of mathematically relating time to position along a parabolic, heliocentric trajectory without prior knowledge of the orbital elements. Around 1770, Lambert added many geometrical approaches to the determination of parabolic orbits in an attempt to reduce the problem to one unknown. Shortly there-

* The amount of northward tilt of this tangent plane is equal to the difference between the geodetic and geocentric latitude of the station and is approximately $11'35.6635'' \sin 2\varphi - 1.1731'' \sin 4\varphi + 0.0026'' \sin 6\varphi$, where φ = the geodetic latitude of the station. See Ref. 5, p. 480.

after, in 1783, LaGrange added the important assumption that all observed positions of an orbiting celestial body (comet, planet, or natural satellite) lie approximately in a single plane. He also introduced a perturbation analysis to determine departures from the assumed plane. LaPlace in 1780 showed a rigorous solution⁶ for an orbit formed from any conic section about the sun, using three positions of the celestial body involved as well as their first and second derivatives.* This method, though elegant, has not come into general use because of the difficulty in determining the position derivatives to sufficient accuracies.

Olbers in 1797, analytically building upon Lambert's geometry, developed a straightforward method for the calculation of parabolic orbits. The method, though not as rigorous as that of LaPlace, is sufficiently accurate and convenient to be used with little change to the present day.

To Gauss belongs the credit of developing the first coordinated approach and practical solution for the case of elliptical orbits.⁷ On January 1, 1801, in searching for possible planets existing in the gap between Mars and Jupiter, the astronomer Piazzi discovered a minor planet later named Ceres. It was observed for a month before it was lost in the sun's glare. A crude projection of its motion, as well as that of the sun, indicated it would not be visible again until about October of that year. Since Ceres shines at only eighth magnitude and appears starlike through any telescope, the search for this minor planet would have been extremely difficult in those days without a good prediction of its position. Within months, Gauss developed a method for determining an elliptical orbit from a sparse bit of "angle-only" data. In his method he reduced the problem to that of two unknowns instead of one for convenience. A resulting eighth-order equation involving range and similar to that produced earlier by LaGrange was solved by Gauss by trigonometric means. The Gaussian method is essentially one of successive approximations which rapidly converge to the true range of a celestial body by appropriate iterations. With the exception of the trigonometric form of the solution that has become unnecessary in the days of high-speed digital computers, the method of Gauss has survived with little modification to the present day. The only significant computational change in the technique occurred in 1851 when Encke recast the method in Cartesian form and concerned himself with rectilinear coordinates rather than the six orbital elements of Gauss. Even now, when LaGrange's planetary equations may be numerically integrated^{9,10,11} and data selected to obtain orbital elements in many chosen forms, the simple form of Gauss' solution is amenable to the addition of closed-form perturbations in an

* For a modern summary of LaPlace's method see Ref. 8, p. 168, or Ref. 9, p. 40.

iterative and rapidly converging mode of machine computation which is most conservative of time.

5.2 General Procedure

The adaptation of Gauss' method in the present paper will make use of the following data and assumptions:

(i) At least three azimuth-elevation predictions of the satellite correlated to time are available.

(ii) The ground station position is known (in terms of latitude, longitude, and height above mean sea level).

(iii) The ground station is not in the orbit plane for any of the data used.

(iv) The instantaneous orbit plane contains the center of mass of the earth.

The computational procedures to follow are now summarized. The first approximation of the range to the satellite is calculated using the above information, a refinement made to this first range estimate, and an initial set of orbital elements developed. Next, perturbations are calculated from the elements, the range is again refined, and so forth until no subsequent change in computed range values is observed to a chosen number of places or until this loop has been traversed a specific number of times. A final calculation of elements is then made and the resulting computed satellite trajectory compared to the measured one as a final test of the validity of the elements. Fig. 3 illustrates this procedure by a flow chart.

5.3 Synthesizing the Range

We begin the detailed synthesis of the range* by referring to Fig. 4, which illustrates an orbit plane about the earth along with three positions of the satellite assumed in that plane. A geocentric right-handed Cartesian coordinate system is also shown, with the z axis coinciding with the earth's spin axis, the x axis pointing toward the vernal equinox† and the y axis orthogonal to these. In this system, which does not rotate with the earth, the orbit plane may be expressed as

$$ax + by + cz = 0. \quad (1)$$

Since the three observed positions of the satellite are initially assumed to lie in this plane, each of these points must satisfy (1) and we may write

* See also Ch. 5 of Ref. 9, Ref. 12, p. 97, and Ref. 8, p. 175. For purely circular orbits see Ref. 13, p. 141.

† See p. 37 of Ref. 14.

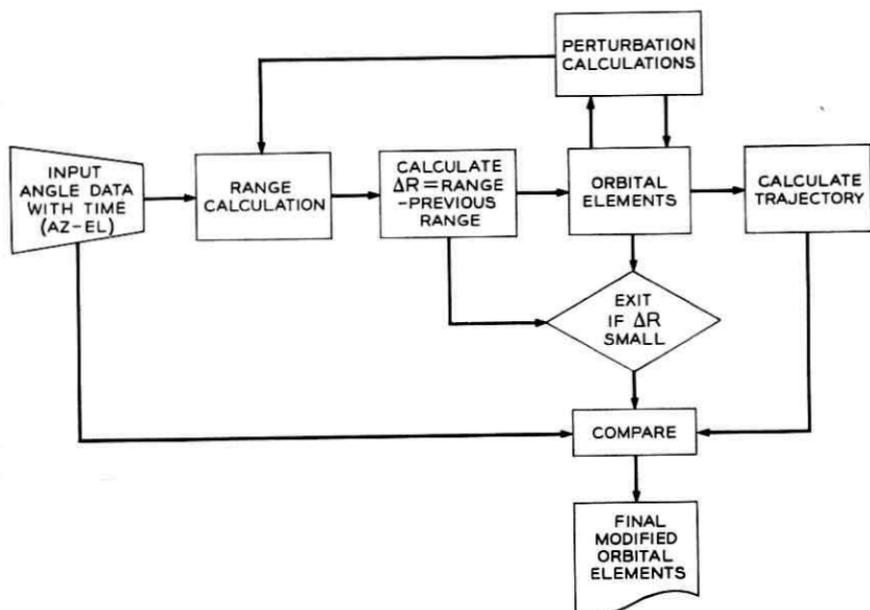


Fig. 3 — Flow chart for MOE generation.

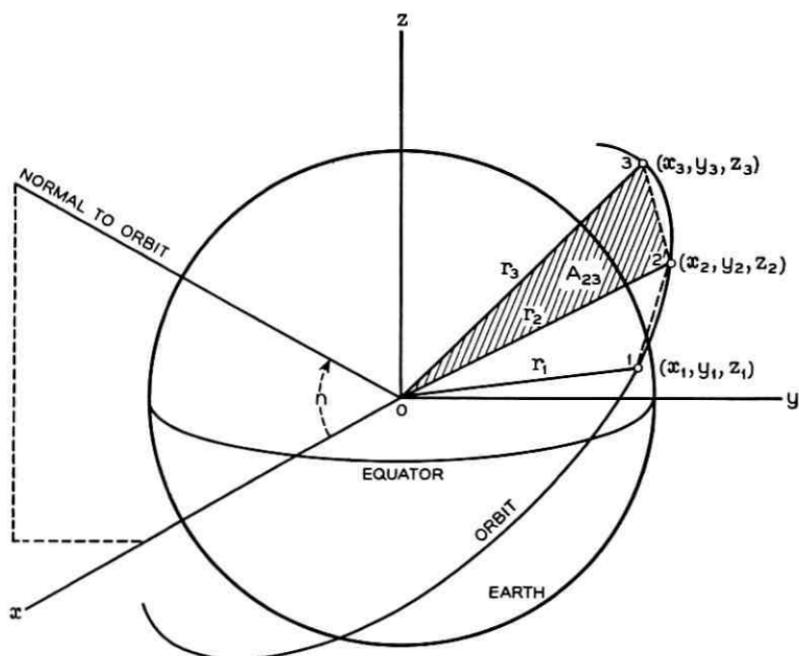


Fig. 4 — The orbit and Gaussian triangles.

$$ax_1 + by_1 + cz_1 = 0 \quad (2)$$

$$ax_2 + by_2 + cz_2 = 0 \quad (3)$$

$$ax_3 + by_3 + cz_3 = 0 \quad (4)$$

from which the determinant of the coefficients becomes

$$\Delta = \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix} = 0 \quad (5)$$

permitting (2), (3), and (4) to be rewritten as

$$(y_2z_3 - z_2y_3)x_1 - (y_1z_3 - z_1y_3)x_2 + (y_1z_2 - z_1y_2)x_3 = 0 \quad (6)$$

$$(z_2x_3 - x_2z_3)y_1 - (z_1x_3 - x_1z_3)y_2 + (z_1x_2 - x_1z_2)y_3 = 0 \quad (7)$$

$$(x_2y_3 - y_2x_3)z_1 - (x_1y_3 - y_1x_3)z_2 + (x_1y_2 - y_1x_2)z_3 = 0. \quad (8)$$

In Fig. 4 twice the area (A_{23}) of the triangle 023 may be related to the coefficient of x_1 through the angle (n) between a normal to the orbit plane and the x axis. The following equation expresses this and is derived in Appendix A:

$$y_2z_3 - z_2y_3 = A_{23} \cos n. \quad (9)$$

Similar equations may be obtained in like manner for triangles 012 and 013. When these are substituted into (6), (7) and (8) one obtains

$$A_{23}x_1 - A_{13}x_2 + A_{12}x_3 = 0 \quad (10)$$

$$A_{23}y_1 - A_{13}y_2 + A_{12}y_3 = 0 \quad (11)$$

$$A_{23}z_1 - A_{13}z_2 + A_{12}z_3 = 0. \quad (12)$$

Normalizing to A_{13} , (10), (11) and (12) become

$$m_1x_1 - x_2 + m_3x_3 = 0 \quad (13)$$

$$m_1y_1 - y_2 + m_3y_3 = 0 \quad (14)$$

$$m_1z_1 - z_2 + m_3z_3 = 0 \quad (15)$$

where

$$m_1 = A_{23}/A_{13} \quad (16)$$

and

$$m_3 = A_{12}/A_{13}. \quad (17)$$

Provided m_1 and m_3 are known, we may solve (13), (14) and (15) for

x_i , y_i and z_i ($i = 1, 2, 3$). Alternatively we may express x_i , y_i and z_i in terms of the slant ranges, ρ_i , of the satellite from the ground station and solve for these ranges in terms of m_1 and m_3 . This latter route will be pursued.

Each satellite position measured from the ground station in terms of azimuth and elevation may be expressed in terms of right ascension and declination, knowing time, by standard techniques⁵ (see also Appendix B). The direction cosines¹⁵ of the line of sight from station to satellite then become

$$a_i = \cos \delta_i \cos \alpha_i \quad (18)$$

$$b_i = \cos \delta_i \sin \alpha_i \quad (19)$$

$$c_i = \sin \delta_i \quad (20)$$

where

α_i, δ_i = the topocentric right ascension and declination of the satellite.

Since the satellite is essentially orbiting the mass center of the earth, it will become advantageous to relate any topocentric coordinates of the satellite as measured from the station to the xyz geocentric ones previously specified. To do this we write

$$x_i = a_i \rho_i + X_i \quad (21)$$

$$y_i = b_i \rho_i + Y_i \quad (22)$$

$$z_i = c_i \rho_i + Z_i \quad (23)$$

where

ρ_i = the slant ranges from the station to the satellite
 X_i, Y_i, Z_i = the XYZ coordinates of the station in the geocentric coordinate system of Fig. 4.

For any instant of time the X_i, Y_i, Z_i coordinates of the station may be easily derived by referring to Fig. 5. By inspection we write the direction cosines, which are quite similar to (18), (19) and (20), as

$$X_i = R \cos L \cos \alpha_i(t) \quad (24)$$

$$Y_i = R \cos L \sin \alpha_i(t) \quad (25)$$

$$Z_i = R \sin L \quad (26)$$

where

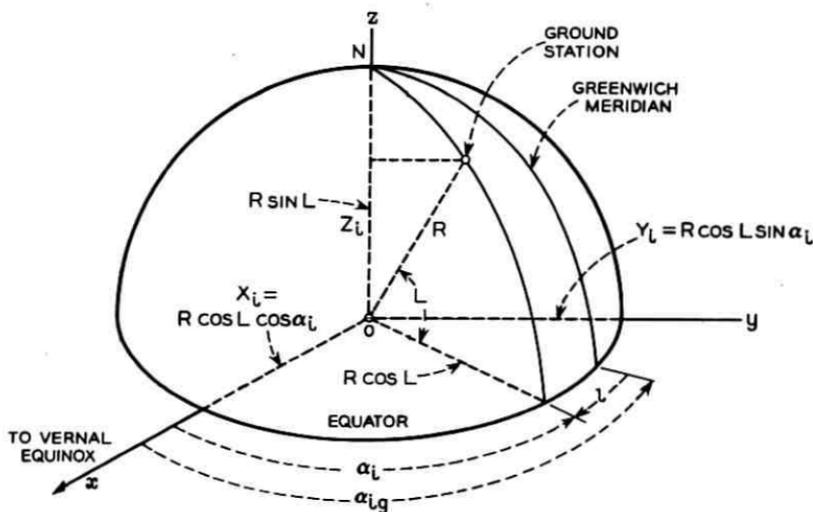


Fig. 5 — Ground station coordinates.

L = station geocentric latitude (see footnote at end of Section IV)

R = distance of station from center of the earth (see Appendix B)

$\alpha_i(t)$ = the right ascension of the station at the instant (t) of satellite observation (local sidereal time at the station).

The local sidereal time* for the above equations may be computed from

$$\alpha_i(t) = \alpha_{ig}(t) - W_1 \quad (27)$$

where

$\alpha_{ig}(t)$ = sidereal time at Greenwich in degrees†

W_1 = west longitude of the station.

A substitution of (21), (22), and (23) into (13), (14), and (15) will accomplish the desired result of expressing x_i , y_i , and z_i in terms of the slant range and topocentric position of the satellite, the station coordinates, and m_1 , m_3 . We obtain

$$a_1 m_1 \rho_1 - a_2 \rho_2 + a_3 m_3 \rho_3 = -m_1 X_1 + X_2 - m_3 X_3 \quad (28)$$

* See Ref. 14, p. 41.

† The term $\alpha_{ig}(t)$ may be computed from the standard meridian time for the station by standard techniques.⁵

$$b_1 m_1 \rho_1 - b_2 \rho_2 + b_3 m_3 \rho_3 = -m_1 Y_1 + Y_2 - m_3 Y_3 \quad (29)$$

$$c_1 m_1 \rho_1 - c_2 \rho_2 + c_3 m_3 \rho_3 = -m_1 Z_1 + Z_2 - m_3 Z_3 \quad (30)$$

where the unknowns are m_1 , m_3 and ρ_1 . A solution for ρ_2 may be obtained if $m_1 \rho_1$ and $m_3 \rho_3$ are eliminated from these last three equations. This yields

$$\rho_2 = D/E \quad (31)$$

where

$$D = m_1 F_1 - F_2 + m_3 F_3 \quad (32)$$

$$E = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \quad (33)$$

$$F_1 = \begin{vmatrix} a_1 & X_1 & a_3 \\ b_1 & Y_1 & b_3 \\ c_1 & Z_1 & c_3 \end{vmatrix} \quad (34)$$

$$F_2 = \begin{vmatrix} a_1 & X_2 & a_3 \\ b_1 & Y_2 & b_3 \\ c_1 & Z_2 & c_3 \end{vmatrix} \quad (35)$$

$$F_3 = \begin{vmatrix} a_1 & X_3 & a_3 \\ b_1 & Y_3 & b_3 \\ c_1 & Z_3 & c_3 \end{vmatrix} \quad (36)$$

Examining (31) through (36) makes it obvious that ρ_2 is determined as soon as m_1 and m_3 are expressed by some separate consideration.

The one piece of data which has not been worked into the analysis so far is the time of flight of the satellite from point 1 to 2 to 3 (see Fig. 4). This information added to the above equations, developed solely from the geometry of the situation, will permit the solution for the range.

In Appendix C, it is shown that m_1 and m_3 are not only ratios of triangular areas formed by the radius vectors to the three satellite positions and their respective chords, but also functions of time* expressible as

$$m_1 = \frac{A_{23}}{A_{13}} \quad (37)$$

$$= \frac{\tau_1}{\tau_2} \left[1 + \frac{\tau_3(\tau_2 + \tau_1)}{6r_2^3} + \frac{\tau_3(\tau_3^3 + \tau_1\tau_3 - \tau_1^2)}{4r_2^4} \frac{dr}{d\tau} + \dots \right]$$

* Similar expressions are found in Ref. 9, p. 52, Eq. (5.3).

$$\begin{aligned}
 m_3 &= \frac{A_{12}}{A_{13}} \\
 &= \frac{\tau_3}{\tau_2} \left[1 + \frac{\tau_1(\tau_2 + \tau_3)}{6r_2^3} - \frac{\tau_1(\tau_1^2 + \tau_1\tau_3 - \tau_3^2)}{4r_2^4} \frac{dr}{d\tau} + \dots \right]
 \end{aligned} \tag{38}$$

where

$$\tau_1 = k(t_3 - t_2) \tag{39}$$

$$\tau_2 = k(t_3 - t_1) \tag{40}$$

$$\tau_3 = k(t_2 - t_1) \tag{41}^*$$

r_2 = geocentric satellite distance at time t_2

t_i = time of i th observation

k = constant for earth satellites†

$$= 0.07436574 \text{ min}^{-1}.$$

In (37) and (38) the geocentric range derivatives at this point are unknown, and so we approximate m_1 and m_3 by considering only the first two terms in the above equations. We shall call these estimates n_1 and n_2 respectively. Returning to (31), the range ρ_2 now may be explicitly expressed in terms of known quantities and the unknown r_2 as

$$\rho_2 \approx (1/E)[n_1 F_1 - F_2 + n_3 F_3] \tag{42}$$

$$\approx (1/E)[(\tau_{12} + p_{12}/r_2^3)F_1 - F_2 + (\tau_{23} + p_{23}/r_2^3)F_3] \tag{42a}$$

where

$$\tau_{12} = \tau_1/\tau_2, \quad \tau_{23} = \tau_3/\tau_2 \tag{43}$$

$$p_{12} = \frac{\tau_1\tau_3(1 + \tau_{12})}{6}, \quad p_{23} = \frac{\tau_1\tau_3(1 + \tau_{23})}{6} \tag{44}$$

The following geometric relationship between ρ_2 and r_2 may be written by inspecting Fig. 6

$$r_2^2 = R^2 - 2R\rho_2 \cos \psi_2 + \rho_2^2 \tag{45}$$

where

ψ_2 = the angle at the station between a line to the earth's center and one to the satellite

= 90° + elevation of satellite above a spherical earth of radius R .

* Using here notation similar to Gibbs solution of Ref. 12, p. 100.

† The constant k is developed in a manner similar to the Gaussian constant for the sun, starting with the expression for the satellite period as $P = 2\pi k(a/r_e)^{3/2}$, where a = the semimajor axis of the orbit ellipse and r_e = the equatorial radius of the earth.

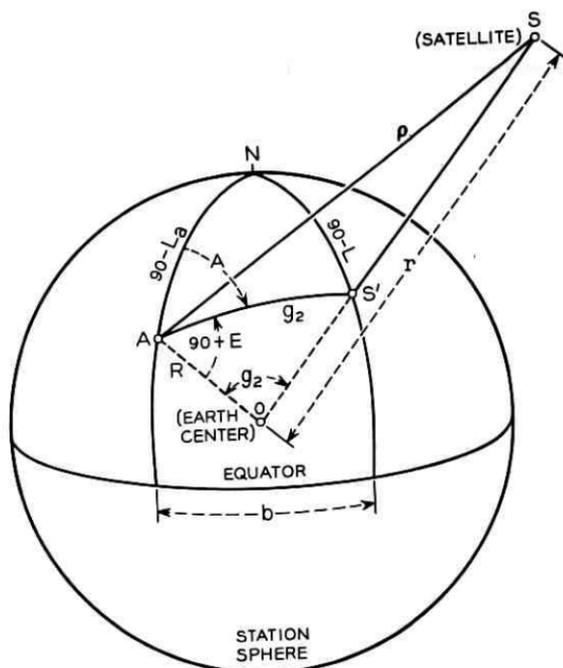


Fig. 6—The station sphere and subsatellite points.

Equations (42) and (45) in combination permit a solution for ρ_2 and r_2 . If one actually proceeds to substitute (42) into (45), an eighth-order equation similar to that derived by LaGrange results. This was later expressed by Gauss as a trigonometric, transcendental equation solved by trial-and-error methods. Rather than pursue this route, it is simpler for machine computations to assume an r_2 value (certainly greater than R_2), calculate a corresponding ρ_2 using (42), enter (45) with the r_2 - ρ_2 pair, determine the departure from equality in this equation, increment the r_2 estimate, and repeat until the departure from equality is sufficiently small. By proper choice of increments, the entire procedure converges rapidly to ρ_2 , r_2 values satisfying the angle-only observations, gravitational theory, and the approximate m_1 , m_3 quantities used. The remaining ρ values easily follow from (28), (29) and (30).

In truncating (37) and (38) after the second term, the range derivatives were omitted in the above analysis. It is possible to return to these equations with these derivatives, which are now crudely known as first-estimate averages over the $t_3 - t_1$ interval. One could then calculate second m_1 , m_3 estimates from (37) and (38), determine new r_i values, compute new geocentric range derivatives, and iterate to an r_i , ρ_i set of

solutions. This is roughly the path followed by LaGrange. By using Gauss' relationship involving certain areas of the orbit ellipse, it is possible to proceed directly to the true values of m_1 and m_3 . This latter course of action, which also allows the approximations to be as close as needed, will now be followed.

In Fig. 4, we define as \bar{y}_i the ratio of the areas (S_i) formed by the radius vectors r_1, r_2, r_3 and the orbit ellipse with the areas (A_i) formed by these same vectors and the chords connecting the three satellite positions. Thus the sector-triangle ratio

$$\bar{y}_1 = S_{23}/A_{23} \quad (46)$$

$$\bar{y}_2 = S_{13}/A_{13} \quad (47)$$

$$\bar{y}_3 = S_{12}/A_{12} . \quad (48)$$

Certainly, in terms of \bar{y}_i , m_1 and m_3 may be defined as [see (37) and (38)]

$$m_1 = \frac{S_{23}\bar{y}_2}{S_{13}\bar{y}_1} \quad (49)$$

$$m_3 = \frac{S_{12}\bar{y}_2}{S_{13}\bar{y}_3} . \quad (50)$$

If we assume pure Keplerian motion, the radius vector from the earth's center to the satellite must always sweep through equal areas in equal times.¹⁴ This permits the direct substitution of time differences for the S factors in (49) and (50) to yield

$$m_1 = \frac{(t_3 - t_2)\bar{y}_2}{(t_3 - t_1)\bar{y}_1} = \frac{\tau_1\bar{y}_2}{\tau_2\bar{y}_1} \quad (51)$$

$$m_3 = \frac{(t_2 - t_1)\bar{y}_2}{(t_3 - t_1)\bar{y}_3} = \frac{\tau_3\bar{y}_2}{\tau_2\bar{y}_3} . \quad (52)$$

It is possible to determine \bar{y}_i knowing the geocentric distance of the satellite and the geocentric angle through which it traveled in the measured time. This technique will be developed specifically for \bar{y}_1 . The remaining variables \bar{y}_2 and \bar{y}_3 are derived in identical fashion. Considering then the second and third observed positions of the satellite, we write

$$\bar{y}_1 = S_{23}/A_{23} = \frac{k(t_3 - t_2) \sqrt{a(1 - e^2)}}{r_2 r_3 \sin 2d} . \quad (53)$$

The numerator of this equation expresses Keplerian motion as the radius vector sweeps area S_{23} while the denominator gives twice the area of the triangle 023. The variable d is simply the difference in the true anomalies

of the satellite at positions 2 and 3. We may shorten the notation by writing

$$\bar{y}_1 = \frac{\tau_1 \sqrt{p}}{r_2 r_3 \sin 2d}. \quad (54)$$

The only unknown in this equation is p , which is a function of the size and shape of the final orbit ellipse. Since estimates of r_2 and r_3 are available at this point, p may be eliminated.⁴ This permits \bar{y}_1 to be expressed by a cubic equation* given below:

$$\bar{y}_1^3 - \bar{y}_1^2 - a\bar{y}_1 - a/9 = 0 \quad (55)$$

where

$$a = \frac{b}{5/6 + c + \epsilon} \quad (56)$$

$$b = \frac{\tau_2}{D_y^3} \quad (57)$$

$$D_y = 2\sqrt{r_2 r_3} \cos d \quad (58)$$

$$c = \frac{r_2 + r_3 - D_y}{2D_y} \quad (59)$$

$$\epsilon = \frac{2}{35}x^2 + \frac{52}{1575}x^3 + \dots \quad (60)$$

$$x = b/\bar{y}_1^2 - c. \quad (61)$$

It is indeed unfortunate that the factor a is ultimately a function of \bar{y}_1 , for if this were not the case an explicit solution for \bar{y}_1 would exist. It turns out that ϵ is sufficiently small that it may be equated to 0 until a first estimate of \bar{y}_1 is found, whereupon it may then be calculated from (61) and (60) and used to derive a better estimate of \bar{y}_1 . The estimate of \bar{y}_1 may be improved as much as desired by traversing this loop a multiplicity of times. Knowing the \bar{y}_i values, m_1 and m_3 are next computed from (51) and (52) and entered in place of the approximate n_1 and n_3 values in (42). The three ranges to the satellite are then calculated as described earlier.

VI. GENERATING THE ORBITAL ELEMENTS

6.1 General Procedure

Three angular positions of a near-earth satellite as observed from a given ground station are sufficient to generate a set of orbital elements.

* See also Ref. 9, p. 56.

The addition of range data permits elements to be computed which are generally more accurate than those developed from angle only information. It will now be shown how these elements may be generated first without regard to any perturbations — as if the satellite orbited a spherical, homogeneous earth. The perturbations arising out of the earth's oblateness will then be computed, and finally their effects will be appropriately combined with the input data to produce the modified orbital elements described in Section III.

6.2 *Generating the Unperturbed Elements*

6.2.1 *Locating the Subsatellite Point.*

Given the azimuth, elevation, and range of a satellite from a specified ground station, we begin by calculating the latitude-longitude coordinates of the so-called subsatellite point. This is the point of intersection of a line, drawn from the satellite to the earth's center, with the surface of what we shall call the station sphere. This sphere is centered on the earth's center and has a radius that equals the geocentric distance of the ground station. Fig. 6 shows this. Here the central angle g_2 , which lies between the station A and the subsatellite point S' , is computed by solving triangle AOS . To do this, we recognize the angle OAS is known from the measured elevation of the satellite reduced to a spherical earth (see Appendix B). Thus

$$\angle OAS = 90^\circ + E \quad (62)$$

where

E = elevation in degrees of the satellite above a plane tangent to the station sphere at the station.

Since the slant range to the satellite, ρ , and the geocentric distance of the station (see Appendix B) are also known, we write

$$\cos(E + g_2) = (R/\rho) \sin g_2 \quad (63)$$

from the law of sines applied to triangle AOS . Solving for g_2 we obtain

$$g_2 = \tan^{-1} \left[\frac{\cos E}{R/\rho + \sin E} \right]. \quad (64)$$

Since g_2 must always be less than 90° , there is no angular ambiguity in this quantity as expressed above.

From the same triangle, the geocentric distance of the satellite becomes

$$r = \frac{\rho \cos E}{\sin g_2}. \quad (65)$$

The latitude and longitude coordinates of the subsatellite point are obtained by solving the spherical triangle ANS' knowing g_2 and the azimuth of the satellite, (A). By the law of cosines, the latitude of S' becomes

$$L = \sin^{-1} [\cos g_2 \sin L_a + \sin g_2 \cos L_a \cos A]. \quad (66)$$

Again, there can be no angular ambiguity since L ranges from -90° to $+90^\circ$.

The longitude of S' is determined from the longitude offset (b) of S' from A . By the law of sines

$$b = \sin^{-1} \left[\frac{\sin g_2 \cos L}{\sin A} \right]. \quad (67)$$

Angle b ranges from -180° to 0° to $+180^\circ$, but the direction (sign) of this longitude offset is directly determinable from the azimuth of the satellite as seen from station A . Thus the subsatellite point has been located in latitude and longitude on the station sphere.* These correspond to the geocentric latitude and longitude of S' measured on the earth.

6.2.2 Computing Orbit Inclination and Ascending Node

Two subsatellite points must be known for the geometrical calculation of the orbit inclination and the ascending node. Consider Fig. 7, where S'_1 and S'_2 represent two known positions of the subsatellite point. The orbit inclination angle i is calculated in the following manner. The great circle arc g_{12} is first computed by standard techniques, knowing the latitude and longitude of S'_1 and S'_2 . It is quite important, however, to remember that the earth has spun through a specific angle during the time that the subsatellite point progressed from S'_1 to S'_2 , and both nodal regression and apsidal advance have occurred. Therefore, S'_2 must be assigned a new longitude suitable to the nonrotating geometry of Fig. 7. This new longitude in degrees is

$$l_2 = l'_2 - \frac{(t_2 - t_1)}{1436} 360 + N \quad (68)$$

where

l'_2 = west longitude of S'_2 from the Greenwich Meridian
 computed from the satellite azimuth, elevation,
 and range

* Certain special cases exist, such as when S' lies on the same meridian as A . These require no additional data and are actually simpler, since then $b = 0$, $g_2 = L - L_a$ and spherical triangle ANS' is not needed.

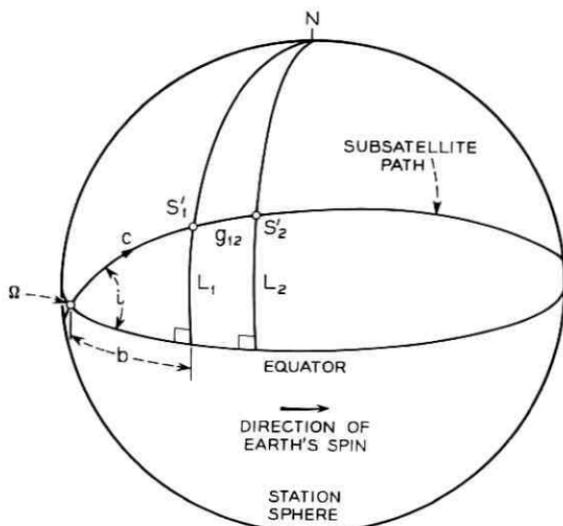


Fig. 7—Geometry for orbit inclination and ascending node calculation.

$t_2 - t_1$ = time in minutes between the occurrence of the two measured subsatellite points

1436 = time for one earth rotation in minutes of mean solar time

N = nodal regression in time $t_2 - t_1$.*

We shall not compute the effects of apsidal advance upon the latitude and longitude of the subsatellite points, as this perturbation will be taken into account in the following section by adjusting the true anomalies.

From Fig. 7, one writes

$$\sin i = \frac{\sin L_1}{\sin c} = \frac{\sin L_2}{\sin (c + g_{12})}. \quad (69)$$

Using the last two terms of (69), one may solve for c to obtain

$$c = \tan^{-1} \left[\frac{\sin g_{12}}{\frac{\sin L_2}{\sin L_1} - \cos g_{12}} \right]. \quad (70)$$

Here $\sin g_{12}$ is always positive, since S_1' and S_2' correspond to satellite

* N is initially set equal to zero until an unperturbed set of elements is generated from which N may be determined: see Section VII.

positions in sight of station *A*. The arc *c*, however, may range from 0° to 180°. If the denominator of (70) is positive, *c* is between 0° and 90°; if negative, *c* is between 90° and 180°.

Knowing *c*, *i* is simply

$$i = \sin^{-1} \left[\frac{\sin L_1}{\sin c} \right] \tag{71}$$

with no ambiguity, since it ranges from 0° to 90° for proper satellites. In similar fashion, *b* of Fig. 7 is

$$b = \cos^{-1} \left[\frac{\cos c}{\cos L_1} \right]. \tag{71a}$$

If $\cos b$ is positive, *b* is between 0° and 90°; if negative, between 90° and 180°. The ascending node is then simply

$$\Omega_1 = l_1 - b \pm 360 n \tag{71b}$$

where

*l*₁ = the W longitude of *S*₁'.

n = an integer chosen to cause Ω_1 to lie between 0° and 360°.

So far only the geometry for *S*₁' and *S*₂' both on the same side of the equator has been considered. If they are on opposite sides, then the only change is in (70), where the - sign becomes a + sign.

6.2.3 Computing the True Anomaly along with Orbit Size and Shape.

If the latitude and longitude positions of the three subsatellite points are known along with the corresponding geocentric satellite distances, the orbit size and shape may be computed and the satellite's true anomaly determined. To do this, consider Fig. 8, where *r*₁, *r*₂, *r*₃, *v*₂₁' and *v*₃₁' are all known. The values for *r*₁, *r*₂, *r*₃ were obtained by (65), and *v*₂₁, *v*₃₁ are merely the great circle arcs between *S*₂' and *S*₁' and between *S*₃' and *S*₁', respectively, corrected for earth spin and later for regression per (68). It is assumed that perigee moves in a secular manner along the orbit ellipse during the time of travel of the satellite from 1 to 2 to 3. Consequently the true anomaly differences determined geometrically must be corrected as follows

$$v_{21}' = v_{21} + \frac{\Delta P(M_2 - M_1)}{2\pi} \tag{72}$$

$$v_{31}' = v_{31} + \frac{\Delta P(M_3 - M_1)}{2\pi} \tag{72a}$$

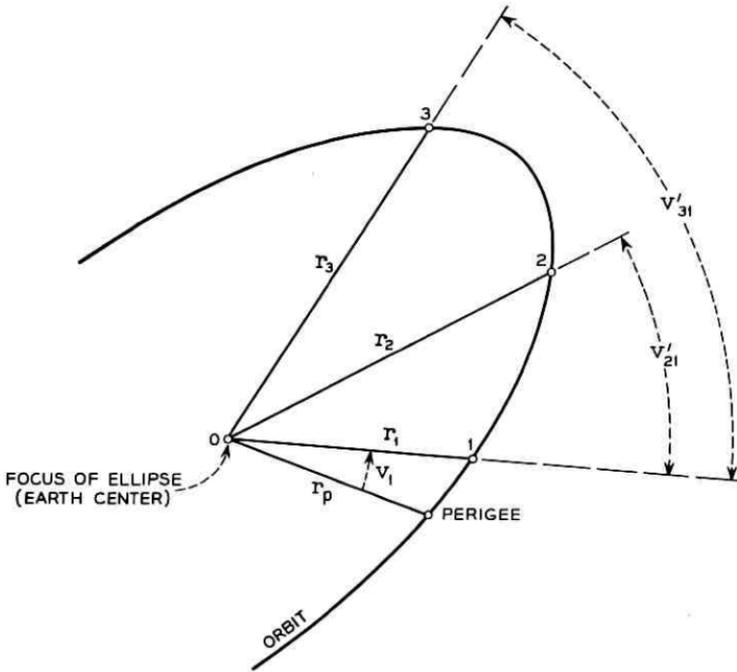


Fig. 8 — True anomaly geometry.

where

v_{21}, v_{23} = the geometric true anomaly differences

ΔP = the apsidal advance in degrees per anomalistic period

M_1, M_2, M_3 = the mean anomalies* of the satellite at 1, 2, and 3 respectively.

Quite clearly the mean anomalies are not known until the position of perigee is determined. Thus, the corrections of (72) and (72a) are not applied to v_{21} and v_{31} until first estimates of the perigee position and of the apsidal advance rate are obtained. At that time a first estimate of the true anomalies of the satellite is known, and this is used to determine a first estimate of the mean anomalies. With these values, (72) and (72a) are used to correct the purely geometrical true anomalies and the loop is traversed until no further significant changes in v_{21}' or v_{31}' are observed.

Let us now set

* See Ref. 9, p. 113.

$$p = a(1 - e^2) \tag{73}$$

where

a = the semimajor axis of the orbit ellipse
 e = eccentricity of the orbit ellipse.

Then for the three satellite positions the polar form of the ellipse* becomes

$$r_1 = \frac{p}{1 + e \cos v_1} \tag{74}$$

$$r_2 = \frac{p}{1 + e \cos (v_1 + v_{21})} \tag{75}$$

$$r_3 = \frac{p}{1 + e \cos (v_1 + v_{31})} \tag{76}$$

where

v_1 = the true anomaly of the satellite at position 1.

It is helpful to rewrite these three equations as

$$e \cos v_1 = (p/r_1) - 1 \tag{77}$$

$$1 + Ae \cos v_1 + Be \sin v_1 = p/r_2 \tag{78}$$

$$1 + Ce \cos v_1 + De \sin v_1 = p/r_3 \tag{79}$$

where A, B, C and D are all known quantities defined as

$$A = \cos v_{21} \tag{80}$$

$$B = -\sin v_{21} \tag{81}$$

$$C = \cos v_{31} \tag{82}$$

$$D = -\sin v_{31} . \tag{83}$$

Solving for p in (77), (78) and (79) yields

$$p = \frac{C - 1 + D(1 - A)/B}{\frac{C}{r_1} - \frac{1}{r_3} + \frac{D}{B} \left(\frac{1}{r_2} - \frac{A}{r_1} \right)} \tag{84}$$

All quantities in the right-hand member of this equation are known, so that p is determined. Returning to (77) and (78), the solution for the true anomaly is

* See Ref. 9, p. 116.

$$v_1 = \tan^{-1} \left[\frac{\sin v_1}{\cos v_1} \right] = \tan^{-1} \left[\frac{\frac{1}{B} \left\{ \frac{p}{r_2} - 1 - A \left(\frac{p}{r_2} - 1 \right) \right\}}{\frac{p}{r_1} - 1} \right] \quad (85)$$

unless $(p/r_1) - 1$ is zero, which occurs for a circular orbit. In this case, the solution for v_1 is of little importance and may be arbitrarily set to 90° .

In general, the true anomaly ranges from 0 to 360° , but all quadrant ambiguities in (85) may be resolved by observing the algebraic sign of $\sin v_1$ and $\cos v_1$, expressed by the numerator and denominator of the final fraction in (85).

With v_1 known, the eccentricity from (77) is simply

$$e = \frac{\frac{p}{r_1} - 1}{\cos v_1} \quad (86)$$

The radius of perigee from (73) is

$$r_p = p \left(\frac{1 - e}{1 - e^2} \right) \quad (87)$$

6.2.4 Reduction to Epoch

One of the features of the modified orbital elements is that they are quoted for the epoch at which the satellite passes through perigee. The node and perigee positions as well as the true anomaly of the satellite determined in the preceding sections were computed for the time of the first data point. They must now be reduced to the time of perigee passage.

During the time the satellite moved from perigee to the first data point, the earth rotated, the orbit node regressed, and the perigee moved. The angular velocity of the earth is known and the node and perigee motions are secular functions of i , e , and r_p , as will be shown later in Section VII. Consequently it is possible by calculation to "move the satellite" back to perigee, taking into account these other motions as well. To do this we shall assume Keplerian motion along an ellipse whose perigee moves in time. If we knew the true anomaly of the satellite at position 1, referenced to the moving perigee, we could determine the time required for the satellite to orbit from perigee to data point number 1. This true anomaly may be obtained in an iterative manner using

$$v_1' = v_1 + \Delta P (M_1/2\pi) \quad (88)$$

where

- ΔP = apsidal advance in degrees per anomalistic period
 M_1 = the mean anomaly of the satellite at data point 1
 v_1 = the geometrical true anomaly of satellite when at point 1 (from Section 6.2.3).

A first estimate of M_1 is obtained from v_1 using Kepler's equation. This permits a first estimate of v_1' by using (88), which in turn allows a second estimate of M_1 based on the first v_1' estimate. The procedure cycles until further changes in v_1' are insignificant. The final value of M_1 corresponding to the final v_1' is used to determine the time of flight of the satellite from perigee to the first data point, which is

$$t_b = (M_1/2\pi)T_a \quad (89)$$

where

T_a = the anomalistic period in minutes.

The elements are back dated by the amount t_b to produce modified orbital elements. In these the ascending node is simply

$$\Omega = \Omega_1 + \Delta\Omega t_b \quad (90)$$

where

- Ω_1 = the west longitude of the node at data point no. 1
 $\Delta\Omega$ = the regressional rate in degrees per minute.

Similarly, the argument of perigee is

$$P = P_1 + \Delta P(M_1/2\pi) \quad (91)$$

where

- P_1 = the perigee position at data point no. 1
 ΔP = the apsidal advance rate in degrees per anomalistic period.

6.2.5 The Anomalistic Period

The anomalistic period may be computed from the mean anomalies at the data points (after these are known from the first estimate of the orbital elements) and the times of flight through these points. For example, one may write

$$T_a = \frac{2\pi(t_2 - t_1)}{M_2 - M_1} \quad (92)$$

or

$$T_a = \frac{2\pi(t_3 - t_1)}{M_3 - M_1} \quad (93)$$

or

$$T_a = \frac{2\pi(t_3 - t_2)}{M_3 - M_2} \quad (94)$$

Equally well, T_a may be computed from the finally determined values of i , e , and r_p by methods outlined in Section VII.

6.3 Adding the Perturbations to the Orbital Elements

In the derivation of the modified orbital elements, perturbations seem required before sufficient orbital data are available for their computation. The general technique is to assume all perturbations zero until a first set of elements has been calculated for epoch t_1 , then to supply these perturbations to produce a second set of elements, and to continue this iteration until further element changes are insignificant. Thus the N of (68), the ΔP of (72), (72a), and (88), and the T_a of (89) follow this route.

For orbits of the TELSTAR satellites which have had eccentricities of up to 0.4 and r_p values as low as 4556 statute miles, convergence by the above method occurs within three iterations. Four iterations are normally required for eccentricities up to about 0.8.

VII. THE SECULAR PERTURBATIONS

7.1 Perturbation Theory and Formulae

The perturbations to be outlined here are secular and produced only by the earth's oblateness. The gravitational field therefore will be independent of the longitude and symmetrical about the earth's equator. Its potential* may be expressed as

$$U = \frac{GM}{r} \left[1 - \sum_{n=1}^{\infty} J_n \left(\frac{R_E}{r} \right)^n P_n(\cos L_c) \right] \quad (95)$$

where

J_n = constants determined in the first instance by observation of many satellites

* See Ref. 16, p. 4.

R_E = the equatorial radius of the earth = 3963.347 statute mi.

r = the geocentric satellite distance

$P_n(\cos L_c)$ = the Legendre polynomial of order n

L_c = the geocentric colatitude of the satellite.

Writing (95) using only the J_2 terms, we have

$$U = \frac{GM}{r} \left[1 + \frac{1}{2} J_2 \left(\frac{R_E}{r} \right)^2 (1 - 3 \cos^2 L_c) \right] \quad (96)^*$$

where

GM = Newton's universal gravitational constant times the mass of the earth^{16,17}

= $(3.44280 \pm 0.000026) \times 10^8$ statute miles³/minutes²

J_2 = $\frac{2}{3} \times$ Jeffreys' J (see Refs. 18 and 19)

= 1.0827×10^{-3} .

Equations of motion for a satellite orbiting this oblate earth are found by setting the spherical polar components of acceleration equal to

$$\frac{\partial U}{\partial r}, \quad \frac{1}{r} \frac{\partial U}{\partial L_c}, \quad \text{and} \quad \frac{1}{r \sin L_c} \frac{\partial U}{\partial \varphi},$$

where r , L_c , and φ are the spherical coordinates of the satellite. After a bit of work, the following secular rates of the orbital elements are produced. For the secular nodal regression† in radians per nodal period we have

$$\begin{aligned} \Delta\Omega &= -3\pi J_2 \left(\frac{R_E}{p} \right)^2 \cos i + OJ^2 \\ &= -0.16028 \frac{\cos i}{p^2} 10^6 \end{aligned} \quad (97)$$

where

* $J_1 = 0$ since the orbit plane is assumed to include the center of mass of the earth.

† Many authors have derived perturbation equations for orbital elements and expressed them in numerous ways, quite often with great elegance and detail. It may seem, therefore, that (97) is unduly simple. This analysis, however, is confined to secular rates only. It can be shown that contributions to these rates from the third through the sixth zonal harmonics of the earth's potential function (as well as from the J_2^2 terms) are about three orders of magnitude less than the J_2 contribution given by (97). They are, therefore, omitted in this analysis where long-term predictions of pointing angles need only be accurate to within 0.05° . Secular contributions to the orbit eccentricity and inclination are also of order J_2^2 and omitted. (See Ref. 21 for expressions of higher-order terms.)

$$p = a(1 - e^2)$$

$$OJ^2 = \text{terms of the order of } J^2 \text{ which are omitted.}$$

The secular apsidal advance rate becomes

$$\Delta\omega = -\frac{3}{2}\pi J_2 \left(\frac{R_E}{p}\right)^2 [1 - 5\cos^2 i + Oe^2]$$

$$= 0.08014 \frac{(5\cos^2 i - 1)}{p^2} 10^6 \quad (98)$$

in radians per nodal period. Finally, the nodal period which expresses the energy in the ellipse is

$$T_n = \frac{2\pi \left(\frac{a^3}{GM}\right)^{\frac{1}{2}} \left[1 - \frac{3}{4}J_2 \left(\frac{R_E}{p}\right)^2 (5\cos^2 i - 1)(1 - e^2)^{\frac{3}{2}}\right]}{1 + \frac{3}{2}J_2 \frac{R_E^2}{a^2(1 - e^2)^3}} \quad (99)$$

where

$$2\pi/(GM)^{\frac{1}{2}} = 3.38629354 \times 10^{-4} \text{ minute/statute miles}^{3/2}.$$

The prime sweep interval of the modified orbital elements is related to the true nodal regression by the following expressions

$$\Delta\Omega' = \frac{\Delta\Omega \times 82505.92}{T_n} \quad \text{deg/day} \quad (99a)$$

$$\text{PSI} = 1440 - \frac{4(\Delta\Omega' + 0.9856)}{1 + \frac{\Delta\Omega' + 0.9856}{360}} \quad \text{minutes.} \quad (99b)$$

The rates for the modified orbital elements are not complete until the anomalistic period has been derived. This may be accomplished knowing T_n and $\Delta\omega$. The procedure is outlined below.

7.2 Determining the Anomalistic Period from the Nodal Period

We have defined the anomalistic period as the time required for the satellite to move from perigee to the next perigee. The two perigee positions are shown as points p_1 and p_2 in Fig. 9, where it is seen that the angular motion of the satellite in the time T_a may exceed 360° . During this time the perigee itself is in motion and its rate is expressed by (98).

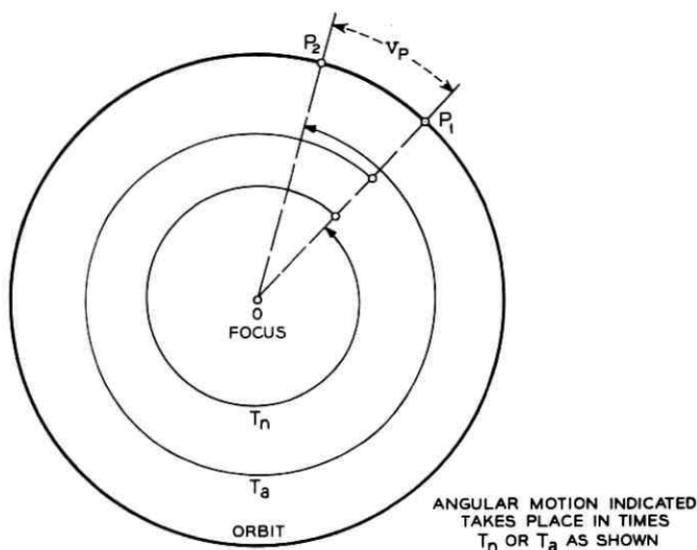


Fig. 9 — Relationship between T_n and T_a .

Thus the anomalistic period exceeds* the nodal period by an amount t so that

$$T_a = T_n + t \quad (100)$$

or

$$T_a = (1 + f)T_n \quad (101)$$

$$= \left(1 + \frac{M}{2\pi}\right) T_n \quad (102)$$

where

t = the time for the satellite to move from p_1 to p_2

f = the fraction of T_n represented by t

M = the mean anomaly corresponding to f .

We may also express the motion of perigee in time T_a as the angle

$$v_p = \Delta\omega T_n/T_a. \quad (103)$$

If T_a is initially set equal to T_n , then (103) represents a first estimate of v_p from which a first estimate of f may be computed from

* If $\Delta\omega$ is negative, T_n exceeds T_a , but the conversion equations are still valid.

$$E = 2 \tan^{-1} \left[\left(\frac{1-e}{1+e} \right)^{\frac{1}{2}} \tan \frac{v_p}{2} \right] \quad (104)$$

and

$$M = E - e \sin E. \quad (105)$$

By (102) a first anomalistic period estimate is obtained which can yield a second estimate of v_p by (103), and the process is iterated until successive changes in T_a are sufficiently small. Since (102) through (105) contain essentially two unknowns, v_p and T_a , closed-form approximate solutions are possible by using truncated series of expansions of the tangent and sine functions. The iterative method, however, has the advantage that the accuracy is limited only by the number of times that the user is willing to traverse the loop.

Knowing the anomalistic period, it is straightforward to express the nodal regression and apsidal advance in degrees per anomalistic period suitable for use in the modified orbital elements.

VIII. ROUND-THE-WORLD ORBIT PREDICTIONS

So far, techniques have been presented for developing orbital elements suitable for predicting future satellite pointing angles. These predictions suffer little error when referred to the ground station which took the original data. However, if predictions are made for other stations, the errors can be greater due simply to the fact that the elements were derived from only a small portion of the orbit ellipse — a fact which makes it difficult to assess the size and shape of the ellipse. The effect will be most pronounced if the latitude of the second station differs greatly from that of the first* as, for example, the case of elements generated from Andover, Maine, trajectories and used for Woomera, Australia, predictions. In previous works this problem is simply solved by obtaining data from stations spaced around the world and using it in orbit generation. In this paper it is desired to confine the data taken to one station and yet predict for remote stations located anywhere. This is accomplished by recognizing that, during the course of a day, a single base station sees a good portion of the orbital ellipse on successive passes. For each such pass a set of elements may be generated. Each set will about equally well describe the orbit over the base station, but can diverge in predictions for stations at different latitudes as illustrated by

* It should be recognized that stations at different longitudes (but at equal latitudes) are not to be considered unique, since they essentially observe the same portions of the orbit at different times of day.

exaggeration in Fig. 10. This is due principally to estimating the size and shape of the orbit ellipse from data which necessarily contain some measurement error. One could improve the situation at remote stations by placing an ellipse of best fit through all the essentially different base station trajectories and thereby gain a better estimate of the eccentricity and radius of perigee in the orbit. If base station predictions made from elements that fit each individual pass are excellent, one could equally well fit an ellipse by a least-squares method to the predicted rather than the measured base station data, and still improve the situation at remote stations. As a matter of fact, proceeding from good base station predictions tends to smooth the input data by bringing them into accord with the motions permitted by gravitational theory and this, in a sense, filters some of the observation noise.

The technique for this round-the-world fit proceeds in detail as follows. Modified orbital elements are used to predict the true anomaly and the

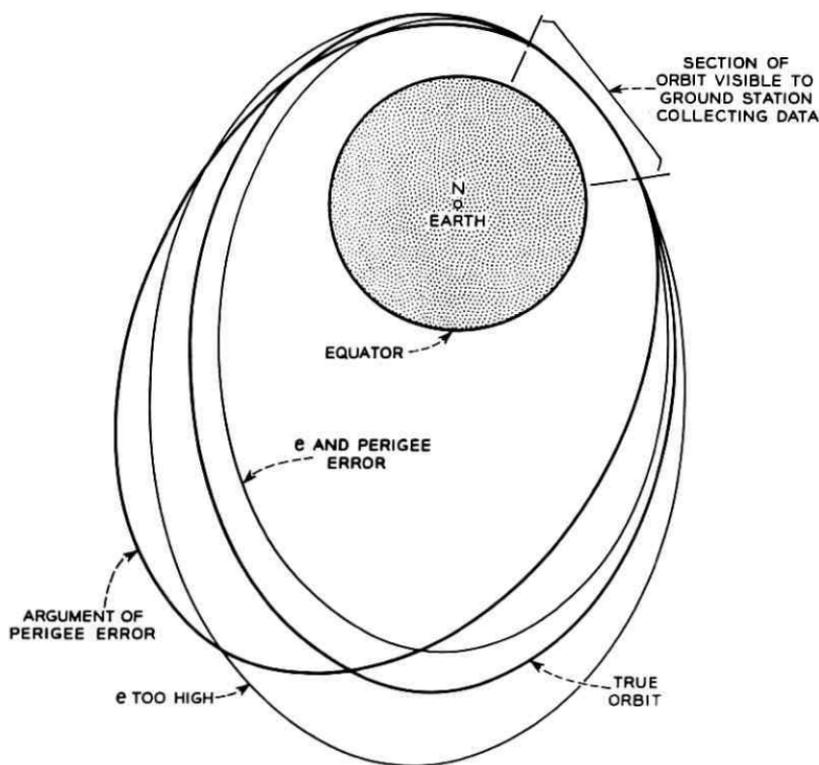


Fig. 10—Projections of orbits on equatorial plane.

corresponding geocentric distance of the satellite for a set of passes seen from the base station. The polar equation of the ellipse is then written as follows:

$$\frac{1}{r} = \frac{1}{a(1 - e^2)} + \frac{e \cos v}{a(1 - e^2)} \quad (106)$$

$$= B_0 + B_1 \cos v. \quad (107)$$

The predicted true anomaly-geocentric distance pairs are then used to determine B_0 and B_1 of (107) by standard least-squares methods. It is important to note that the true anomaly v of (107) must always be referred to the moving orbit perigee as was done in Section 6.3.

Providing that e and r_p do not seriously change their magnitude during the set of passes being considered (that is, no great orbit energy loss), we may determine these quantities from the least-squares B_0 and B_1 values to be

$$e = B_1/B_0 \quad (108)$$

$$r_p = \frac{1 - e}{B_0(1 - e^2)}. \quad (109)$$

While the determination of orbit size and shape for round-the-world orbit predictions may be carried out by least-squares methods, the node position and argument of perigee are often treated differently, since they are not determined to the same degree of accuracy by each pass of the satellite over the base station. Pass selection is often a good procedure to follow for these parameters. For example, data from a pass that contains perigee are generally considered much more likely to produce a more nearly accurate measure of the perigee argument than those from a pass which does not. The same is true of the ascending node.

IX. IMPROVING THE ELEMENT RATES BY DIRECT MEASURE

Quite obviously, if the modified orbital elements are known at two separate epochs, the secular rates may be computed directly. Let us assume that the elements are correct except for the rates of nodal regression, anomalistic period, and apsidal advance.

9.1 Nodal Regression

The west longitude of the ascending node at the earlier epoch (Ω_1) is related to its position at the later epoch (Ω_2) by

$$\Omega_2 + 360A - \Omega_1 = \frac{360L_{12}}{\text{PSI}} \text{ degrees} \quad (110)$$

where

- t_{12} = time interval in minutes between the two epochs
 PSI = the nodal regression expressed as the prime sweep interval (1440 - M minutes)
 A = an integer which may have any positive, negative or zero value.

Examination of (110) reveals ambiguities in multiples of 360° in the determination of the prime sweep interval, since without some knowledge of its magnitude, one cannot be sure if the node has regressed more than one revolution to its Ω_2 position. The theoretical value of the prime sweep interval is available, however, by the techniques of Section VII. Thus the ambiguity is avoided unless t_2 is extremely large, which allows even small variations in PSI to promote as much as one revolution difference in nodal regression. Solving (110) for the prime sweep interval, we have

$$\text{PSI} = \frac{360t_{12}}{\Omega_2 - \Omega_1 + 360A} \text{ minutes.} \quad (111)$$

9.2 Anomalistic Period

From the epochs of the two sets of modified orbital elements on hand, a single anomalistic period, having zero rate of change, may be computed. First, the whole number of satellite passages through perigee in the time t_{12} between epochs is expressed as

$$A_n = \text{INTF } (t_{12}/T_{at}). \quad (112)$$

Then a refined estimate of the anomalistic period becomes

$$T_a = t_{12}/A_n \quad (113)$$

where

- T_{at} = theoretical anomalistic period computed from the first set of MOES using (99) through (105)
 INTF = indicates that only the integer portion of the division is to be retained.

Sometimes the anomalistic period of a particular set of elements has been well refined by the above procedure, so that it is quite precise over a short period of time. If, some time later, a second set of elements is determined and shown to fit the orbit well at this later time, then it is often a good procedure to retain the initial anomalistic period and compute a period change to be consistent with the satellite's arrival at perigee

on epoch 2. This incidentally, is the only manner in which a change in period is deduced for the TELSTAR satellites. The procedure is as follows. Calculate A_n using (112) as before. A first estimate of the change in period per period becomes

$$\Delta T_a = \frac{2t_{12} - A_n T_{at}}{A_n^2}. \quad (114)$$

It is possible that the A_n value calculated by (112) is not correct owing to the change in the period during the interval t_{12} . Therefore we compute a time interval t_f as a trial to compare to t_{12}

$$t_f = A_n T_{at} + (A_n^2/2)\Delta T_a \quad (115)$$

and if t_f exceeds t_{12} , we decrease A_n by 1 and repeat the operations of (114) and (115) in an iterative sense until t_f is less than t_{12} . (Of course, if t_f ever equals t_{12} in the process, then the corresponding A_n value is correct and no further iteration is needed.) When t_f is less than t_{12} compute

$$t_{f1} = (A_n + 1)T_{at} + \frac{1}{2}(A_n + 1)^2\Delta T_a \quad (116)$$

and if t_{f1} exceeds t_{12} , the current A_n value is valid. If t_{f1} is equal to t_{12} then A_n should be decreased by one. If t_{f1} is less than t_{12} , increase A_n by one and repeat the process from (114) on. Eventually the procedure converges on an A_n value that either makes t_f equal to or less than t_{12} and t_{f1} greater than t_{12} . The corresponding ΔT_a is the proper one to use.

9.3 Apsidal Advance Rate

The apsidal advance rate may be expressed as

$$\Delta\omega = \frac{\omega_2 - \omega_1 + 360B}{A_n + f} \quad (117)$$

where

$$f = \frac{t_{12} - t_f}{T_{at}} \quad (118)$$

and

B = an integer chosen by comparing $\Delta\omega$ with the theoretical value computed from the first set of elements, as was done with A of (110).

In cases where no ΔT_a is computed, the A_n value of (112) and $f = 0$ should be used in (117).

The advance rate of (117) is expressed in degrees per anomalistic

period, which has direct meaning provided no change in period is calculated. Whenever ΔT_a is not 0, (117) produces the advance rate which when applied to each new period* between epoch 1 and epoch 2 results in the motion of perigee from ω_1 to ω_2 as stated in the elements.

X. A COMPUTER-OPERATOR ENSEMBLE FOR ORBIT GENERATION

Computational techniques have been developed to permit modified orbital elements to be determined initially from the powered flight parameters of the launch vehicle and subsequently from measured satellite trajectories over a ground station using techniques previously described in this paper. These have been augmented to enable the elements to be refined so as to make them valid for several months and to monitor their validity on a routine basis. The techniques have been incorporated into five computer programs which have been routinely used, in the case of the second TELSTAR satellite, to produce elements for antenna drive tapes, pass scheduling, and satellite attitude predictions. It is the purpose of this section to show the means by which the programs are integrated into a computer-operator system.

10.1 *The Computer Programs*

The five computer programs employed for orbit generation, refinement, and monitoring techniques are:

- (i) The MOEGEN G3 program
- (ii) The MOERATE program
- (iii) The REFINE PERIOD program
- (iv) The CORD DOPPLER G2 program
- (v) The MONITOR TAPE ANALYSIS program.

The first four of these were written by the author, while the last²² was basically developed by D. A. Aaronson with certain subroutines by D. A. Ramos. In this section we shall discuss the general function of each program.

10.1.1 *The MOEGEN G3 Program*

The MOEGEN G3 Program generates modified orbital elements for a near-earth satellite given a list of azimuth, elevation, and, if available, range of the satellite as seen from a ground station at specified times. If just these data are used, the resulting MOES are said to be free-fit. If the PSI, apsidal advance, and anomalistic period are known from past data, the MOEGEN program may be set to accept these rates along with trajectory information as before and produce a suitable set of MOES.

* A new period is calculated each time the satellite passes through perigee.

These MOES are said to be forced. Since the anomalistic period cannot be accurately determined from a single pass, this latter mode of operation is most valuable in producing elements of long lasting validity.

In addition to generating elements, the MOEGEN program compares trajectories generated from each set of MOES it produces with the points used in the generation, as well as with any additional submitted points from any of the passes from that ground station. Here the program reports azimuth differences, elevation differences, and great circle arc pointing errors as well as range errors. Since the program produces a set of MOES for every group of three azimuth-elevation-range points, these comparisons aid in the selection of the best set of MOES from the group generated. The actual selection, which will be discussed later, is done by an operator, since typically only a few sets of MOES are produced on each generation.

10.1.2 *The MOERATE Program*

The MOERATE program, as its name suggests, refines the PSI, apsidal advance, and period rates along the lines described in Section IX by examining two sets of elements, usually the ones previously in use and the one just generated. In addition, it compares the rates so established to theoretical rates determined from the previous set of MOES using the gravitational theory given in Section VII. In generating rates, higher-order changes beyond the PSI, apsidal advance, and change in period are neglected. For the TELSTAR satellites this omission has resulted in no serious errors in the orbit.

10.1.3 *The REFINE PERIOD Program*

Because computed satellite trajectories are most sensitive to the value calculated for the anomalistic period, the REFINE PERIOD program offers an alternate means for its determination and is particularly valuable before monitor tapes* are available. The program requires a set of MOES and times of meridian crossings from any specified ground station. From each of these it computes an anomalistic period assuming all other element values to be accurate. If there are no errors in the recorded time for each meridian crossing, in the adopted station coordinates, or in the elements, quite naturally all the computed period values should be identical. Any scatter in the computed periods points up errors in the above quantities. If the meridian crossing times and station locations

* These are magnetic tapes generated at the ground station. They contain, among other data, the measured trajectory of the satellite.

are known to be correct, the scatter acts as a figure of merit for the generated modified orbital elements.

10.1.4 *The CORD DOPPLER G2 Program*

This program accepts either MOES or, in conjunction with a subroutine called MOE, expected burnout parameters for the final powered stage of the satellite launching vehicle. From these it produces an ephemeris for passes over any specified ground station. If burnout quantities are used, the program converts these to MOES, which it prints prior to the ephemeris listing. In addition, this program also produces world maps of the satellite ground trace, drive tapes for both the Andover and Holmdel ground stations, and reports the solar illumination status of the satellite along with a number of communication quantities.

10.1.5 *The MONITOR TAPE ANALYSIS Program*

The monitor tape analysis program written by D. A. Aaronson and described in detail elsewhere²² compares a satellite trajectory as recorded on a compressed or uncompressed monitor tape²² to that produced by a set of MOES. Its output is a printed sheet giving great circle arc pointing errors and slant range errors (provided ranging data are present on the monitor tape). It also indicates received microwave carrier power so that one can ascertain whether or not the Andover horn antenna was indeed pointed toward the satellite. This is, of course, essential to proper assessment of the printed errors.

10.2 *Orbital Element Generation Operational Procedures*

The generation of modified orbital elements for the TELSTAR satellites falls into two categories: the initial generation just after launch and the subsequent generation after monitor tapes have become available. Figs. 11 and 12 show the logic operations for these two cases.

10.2.1 *Initial Element Generation*

The initial elements are determined from the expected burnout parameters* of the final powered stage of the launch vehicle. These parameters are based upon the telemetry from the first two stages. Basic

* For the TELSTAR satellites, these have been supplied by J. W. Timko of Bell Laboratories.

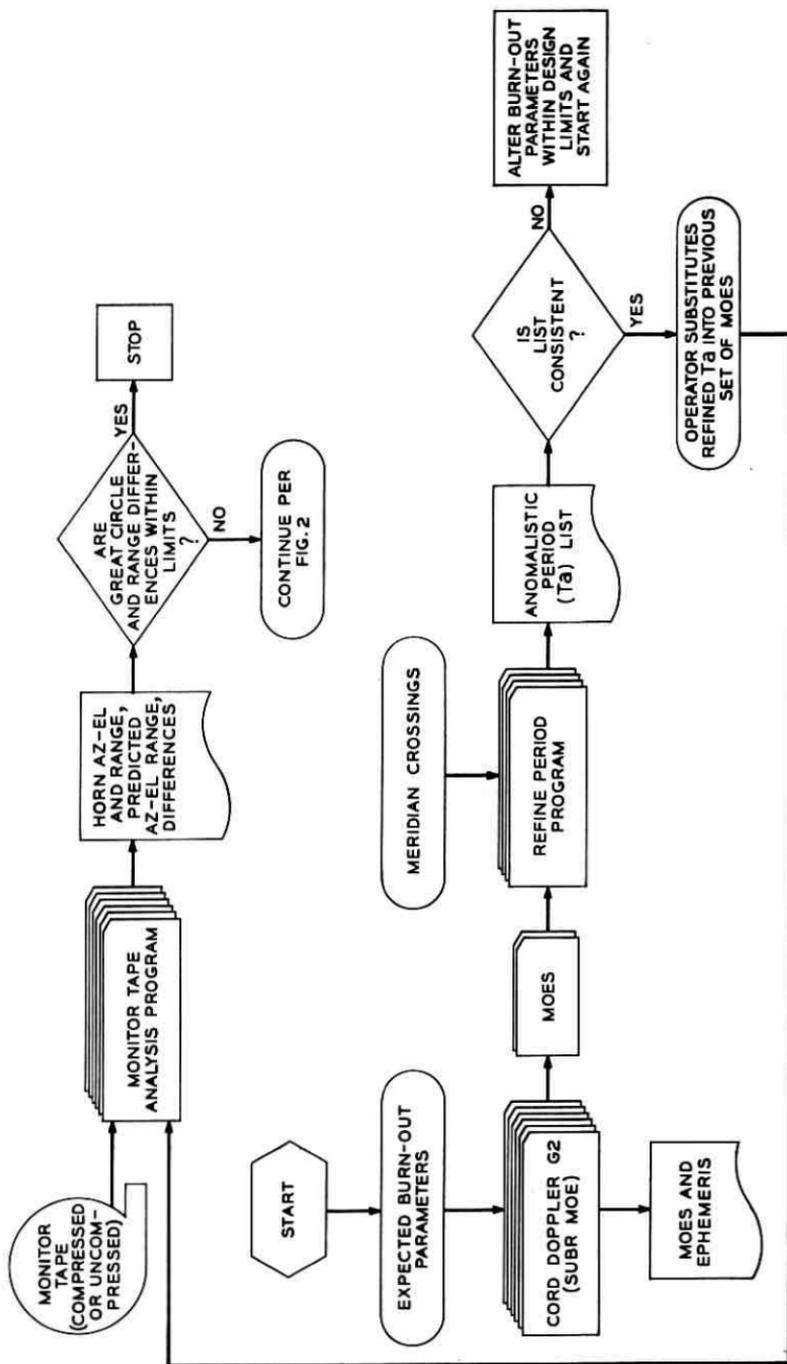


Fig. 11 — Logic blocks for initial modified orbital element generation.

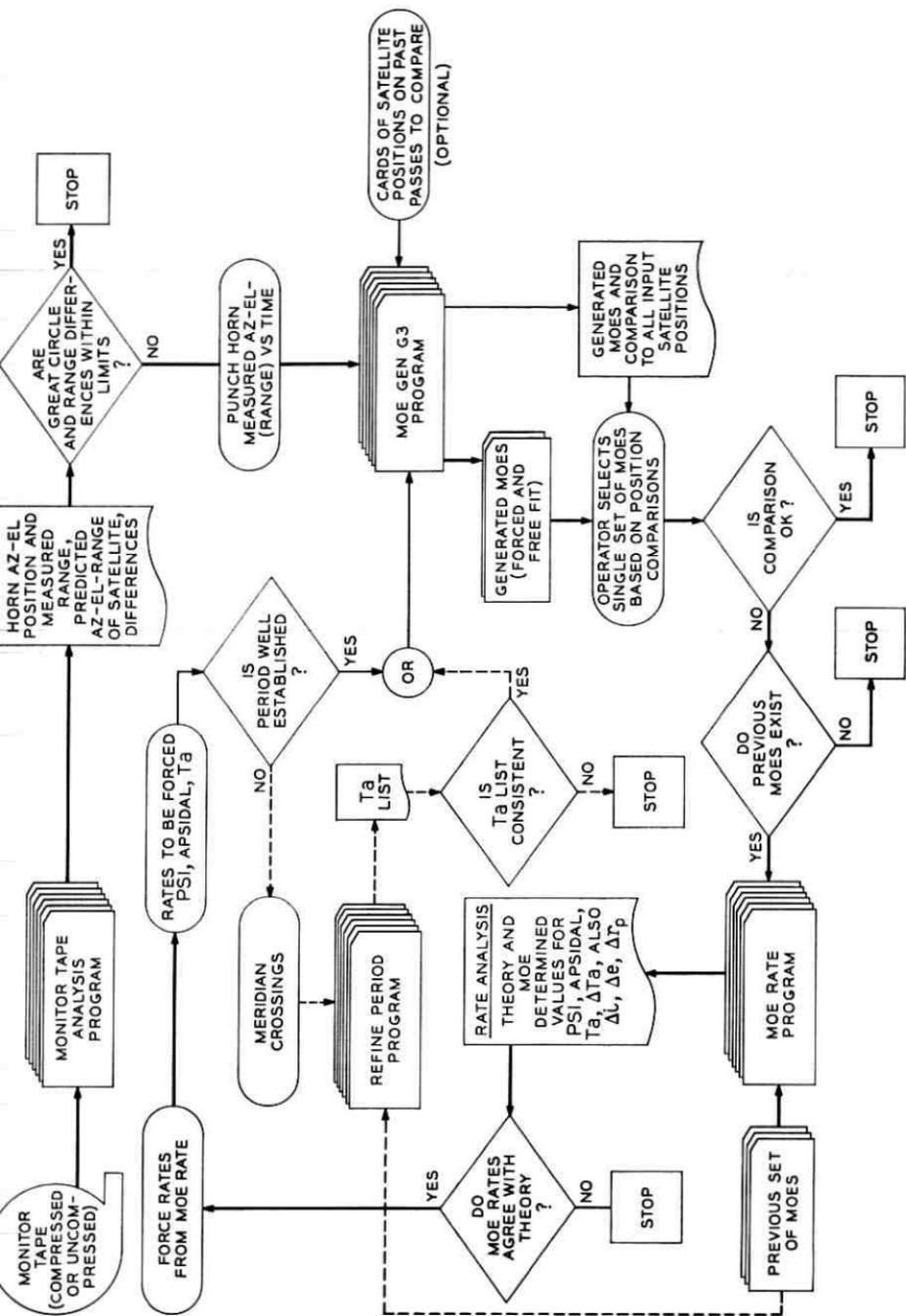


Fig. 12 — Logic blocks for subsequent modified orbital element generation.

to these parameters is the assumption that the third stage performs in a design center manner.

Subroutine MOE of the CORD DOPPLER G2 program accepts the burnout parameters as shown in Fig. 11 and calculates corresponding MOES, which are passed to the main program for ephemeris generation listings.

Modified orbital elements calculated from the burnout parameters are submitted along with subsequent meridian crossings to the REFINE PERIOD program to produce an anomalistic period list as described in Section 10.1.3 (see also Fig. 11). The operator examines this list for consistency. Because the REFINE PERIOD program bases its period calculation on what it assumes as an otherwise correct set of MOES and because the final stage of the launch vehicle, often without telemetry provision, may have performed in a manner different than assumed for orbit generation, it is usually satisfactory to call the period list consistent if the spread of values does not exceed a few tenths of a minute or about 0.1 per cent of the mean anomalistic period for the distribution. If the list is satisfactory according to this criterion, the operator substitutes the mean value of the period into the set of "burnout" MOES and checks minute by minute the predicted trajectory with measured ones by means of the MONITOR TAPE ANALYSIS program as soon as the monitor tape becomes available (see Fig. 11).

Great circle arc errors and range errors reported by the monitor program are scanned by an operator, and if these are within established limits, the element generation process stops and the set of MOES on hand is used for future predictions. If limits are exceeded, the operation proceeds as outlined by Fig. 12, to be discussed later.

Care should be exercised in establishing error criteria for this operator scanning. The first set of MOES, being based on expected burnout parameters, cannot generally produce pointing errors below 0.1° in arc or ± 10 miles in range.

If the anomalistic period list originally generated by the REFINE PERIOD program has too great a spread, an inconsistent set of MOES is generally indicated (see Section 10.1.3). This can occur when the final stage of the launch vehicle departs sufficiently from its design trajectory. It is useful at this time to have on hand a number of MOES generated from burnout parameters with tolerances within the final-stage performance limits. These can be used with the REFINE PERIOD program with the expectation that at least one set will produce an acceptable anomalistic period list. If this procedure is unsuccessful, there

is no recourse but to wait for the monitor tape or sufficient data from tracking stations to establish a correct orbit prediction.

10.2.2 *Subsequent Element Generation*

As soon as monitor tapes become available from Andover, the operation passes into what is termed the subsequent orbital element generation stage. This is outlined in Fig. 12.

As before, monitor tape trajectories are compared to those generated by the current orbital elements. If these comparisons are unsatisfactory (see Section 10.2.1), the horn azimuth, elevation and the measured range of the satellite along with the corresponding times of measurement are fed to the MOEGEN G3 program. Only a free fit (see Section 10.1.1) is made if no previous rate data are available. Both free and forced fits are made if reasonably reliable values for the prime sweep interval, apsidal advance, and anomalistic period are on hand. In the early phases of orbit determination, the REFINE PERIOD program is used to establish an anomalistic period by meridian crossings. After the first workable set of MOES is established by the subsequent generation procedure, the MOERATE program then supplies most of the period values.

In normal operation, the MOEGEN G3 program will compare the input data points to trajectories generated from all output MOES. Optionally, past trajectory data may also be submitted to the program for similar comparisons. An operator scans these comparisons, particularly those concerning the trajectory from which the MOES were derived. The free-fit comparison should have great circle arc differences below 0.05° and range errors below a few miles provided the input trajectory data are consistent with the pointing angles resolved to within 0.02° , as is generally true on the Andover monitor tapes. The forced fit may produce greater errors, which naturally depend on the correctness of the forced data. By these comparisons the operator selects a single set of MOES. Failure to select can occur if input data are excessively noisy or for some reason inconsistent. There is no alternative in this case but to wait for the next monitor tape.

The MOES selected are used directly for future predictions if no previous set other than those from burnout exists. The burnout MOES, being only expected values and having only an approximate epoch, are generally not suitable for submission to the MOERATE program. If a previous set of elements from MOEGEN G3 are available, they are compared to the current set by MOERATE to determine more accurately

the PSI, apsidal advance, and anomalistic period. The operator substitutes these for the earlier rates and again enters MOEGEN G3 using the forced fit mode only. The resulting elements which best fit the present and past trajectories are selected for subsequent use.

XI. TYPICAL EXPERIMENTAL RESULTS

In this section, a typical orbit generation procedure will be illustrated along the lines described in Section 10.2.2 and then a theoretical round-the-world orbit fit will be shown using the techniques of Section VIII. Data for the typical element generation will come from 20 minutes of measurement on each of two passes occurring over Andover on June 30 and July 30. Data from passes on other occasions will be used to determine the accuracy of the predicted trajectories but, for purposes of this illustration, in no way will these latter data be used to refine the orbital elements except as specifically stated.

Fig. 13 shows an output from the MOEGEN G3 program operated in the free-fit mode. Three sets of modified orbital elements were calculated from three data points (azimuth, elevation, range) taken on June 30, 1964. These three sets, shown on lines 7, 8, and 9, differ from each other only in the method of determining the anomalistic period. In the first set, the period is determined from the elements using the perturbation formulations of Section VII. In the second and third sets, the period is determined from the time of flight of the satellite between the data points per Section 6.2.5.

Directly beneath the elements, the program states that the orbit offset is 31.48778° . This is merely an indicator which means that the station is that number of degrees of latitude removed from a point of equal longitude on the orbit plane at the time of the first data point. If this quantity were zero, then the station would be in the orbit plane and this three-point analysis would be invalid.*

The portion of the output enclosed in parentheses displays the comparison of predicted-to-measured trajectories. There are three sets of comparisons, corresponding to the three separate sets of modified orbital elements generated. Here we see that arc and range errors are reasonably small for the June 30 pass (which data were used to determine the elements) but that the fit on June 10, for example, exhibits 5° pointing errors. A survey of the fits clearly indicates that the orbital rates (i.e., prime sweep interval, period, and perigee change) need adjustment.

* Of course the analysis may be rendered invalid if the station is in the orbit plane at the time of *any* of the 3 data points. The first point was reported only as a matter of convenience. In any event the comparison of predicted and measured trajectories described later serves as a final check on the validity of the elements.

Fig. 14 shows the fit of an orbit generated from July 30 data only (azimuth, elevation, range). Here the pointing fit is within 0.02° for that day, but degrades to 0.4° a day later and is grossly inadequate when matched with data from the June 2 trajectories. Again better rate information is needed.

In Fig. 15, we employ the MOERATE program output to determine the orbital rates needed to connect the June 30 and July 30 elements (see Sections IX and 10.1.2). With but a few comments this output should be self-explanatory. First of all, the theoretical values at the bottom of the page are computed from the set of MOES listed at the top left of the figure. The chief purpose for their computation is to gain an estimate of the rates for the actual rate calculations. One cannot compute an average anomalistic period from two given epochs if an estimate of the period is not available. Without such an estimate, the satellite may have made any number of passes through perigee between the two epochs and an infinite array of average periods would be valid solutions. Even with such an estimate, it is sometimes possible to miscalculate the number of times a satellite has passes through perigee if the time between epochs becomes sufficiently large. That is why the number of perigee passes computed by the program is printed out. Fig. 15 (see checked line) has been chosen to show a misestimate of one in the calculation of the perigee passages (= 196) and the attendant error in the average period. By increasing the perigee passes to 197 and performing the indicated division, a new average period of 225.30083 is calculated which is closer to that reported by the first set of MOES and is, of course, a valid solution. Strictly speaking, if this new period value is used, the apsidal advance per T_a should be corrected by multiplying it by $225.30083/226.45032$ to obtain $0.19129^\circ/T_a$. The PSI is unaffected by these operations.

Printed at the very bottom of Fig. 15 is the number of iterations required for the computed PSI and apsidal advance rate to come within 2.1 minutes and 0.2 degrees of the theoretical rates. The program stops after 21 iterations whether or not the convergence is successful.

Since there is no reason to suspect that the anomalistic period associated with the June 30 orbit generation possesses high validity (see Fig. 13), the portion of the MOERATE output dealing with changes in period is ignored and the indicated average anomalistic period is used. From Fig. 15, the computed rates become

$$\begin{aligned} \text{anomalistic period} &= 225.30083 \text{ minutes} \\ \text{apsidal advance rate} &= 0.19129^\circ \text{ per anomalistic period} \\ \text{prime sweep interval} &= \text{one day} - 8.12511 \text{ minutes.} \end{aligned}$$

These rates are then submitted along with the June 30 data to MOE-

MOEGEN G3 - ORBIT GENERATION FROM ANDOVER TRAJECTORIES - L C THOMAS *

RANGE DATA SUPPLIED ON ALL THREE PRINTS. FREE FIT

REFRACTION CORRECTION APPLIED TO INPUT AND OUTPUT SATELLITE ELEVATIONS.

COMPUTED MODIFIED ORBITAL ELEMENTS-

PRINT ITER NO	EPOCH YEAR MO DY HR MINUTE	INCL DEGREES	ASC WEST LONG MINUTES	ARG PERIG DEG	PERIGEE CHG DEG/TA	ANOM PERIOD MIN	PERIOD CHG MIN/PERIOD	ECCEN-TRICITY	RAD PERIGEE STAT MILES
1	1964 6 30 2 53.962	42.76190	219.33549	-8.15271	322.80277	0.19015	225.33698	0.	0.40079
	1964 6 30 2 54.019	219.34985	322.80277	0.19015		225.24235			4567.911
	1964 6 30 2 54.011	219.34777	322.80277	0.19014		225.25511			

ORBIT OFFSET = 31.48778 DEG OF LAT. IF 0 1ST SUBSAT POINT IN ORBIT PLANE.

IYEAR	MO	DY	HR	MN	SEC	PRINTING ANGLE COMPARISON - STATION AT ANDOVER ME									
						M-E-A-S-U-R-E-D	(M-E-A-S-U-R-E-D	(M-E-A-S-U-R-E-D	(M-E-A-S-U-R-E-D	(M-E-A-S-U-R-E-D	(
1964	6	30	5	10	-0	AZ-DEG	EL-DEG	RANGE-MI	AZ-DEG	EL-DEG	RANGE-MI	AZ-DIFF	EL-DIFF	ARC-DIFF	RANGE-DIFF
1964	6	30	5	20	-0	210.36	37.55	7446.59	210.36	37.43	7446.58	-0.000	0.021	0.0232	0.008
1964	6	30	5	30	-0	201.09	31.35	7347.35	201.08	31.32	7347.45	0.007	0.030	0.0328	0.098
1964	6	30	5	40	-0	193.71	23.77	7214.50	193.70	23.72	7214.50	0.010	0.049	0.0514	0.003
1964	6	30	5	50	-0	175.88	25.01	6175.84	174.88	24.87	6146.51	0.997	-11.359	11.3911	29.333
1964	6	30	5	00	-0	175.94	27.96	6186.80	174.87	24.87	6171.00	1.070	-11.139	11.1741	15.800
1964	6	30	5	10	-0	175.97	30.83	6201.62	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	20	-0	175.97	33.70	6216.44	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	30	-0	175.97	36.57	6231.26	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	40	-0	175.97	39.44	6246.08	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	50	-0	175.97	42.31	6260.90	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	00	-0	175.97	45.18	6275.72	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	10	-0	175.97	48.05	6290.54	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	20	-0	175.97	50.92	6305.36	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	30	-0	175.97	53.79	6320.18	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	40	-0	175.97	56.66	6335.00	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	50	-0	175.97	59.53	6349.82	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	00	-0	175.97	62.40	6364.64	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	10	-0	175.97	65.27	6379.46	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	20	-0	175.97	68.14	6394.28	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	30	-0	175.97	71.01	6409.10	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	40	-0	175.97	73.88	6423.92	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	50	-0	175.97	76.75	6438.74	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	00	-0	175.97	79.62	6453.56	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	10	-0	175.97	82.49	6468.38	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	20	-0	175.97	85.36	6483.20	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	30	-0	175.97	88.23	6498.02	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	40	-0	175.97	91.10	6512.84	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	50	-0	175.97	93.97	6527.66	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	00	-0	175.97	96.84	6542.48	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	10	-0	175.97	99.71	6557.30	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	20	-0	175.97	102.58	6572.12	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	30	-0	175.97	105.45	6586.94	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	40	-0	175.97	108.32	6601.76	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	50	-0	175.97	111.19	6616.58	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	00	-0	175.97	114.06	6631.40	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	10	-0	175.97	116.93	6646.22	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	20	-0	175.97	119.80	6661.04	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	30	-0	175.97	122.67	6675.86	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	40	-0	175.97	125.54	6690.68	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	50	-0	175.97	128.41	6705.50	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	00	-0	175.97	131.28	6720.32	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	10	-0	175.97	134.15	6735.14	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	20	-0	175.97	137.02	6749.96	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	30	-0	175.97	139.89	6764.78	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	40	-0	175.97	142.76	6779.60	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	50	-0	175.97	145.63	6794.42	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	00	-0	175.97	148.50	6809.24	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	10	-0	175.97	151.37	6824.06	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	20	-0	175.97	154.24	6838.88	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	30	-0	175.97	157.11	6853.70	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	40	-0	175.97	160.00	6868.52	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	50	-0	175.97	162.87	6883.34	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	00	-0	175.97	165.74	6898.16	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	10	-0	175.97	168.61	6912.98	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	20	-0	175.97	171.48	6927.80	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	30	-0	175.97	174.35	6942.62	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	40	-0	175.97	177.22	6957.44	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	50	-0	175.97	180.09	6972.26	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	00	-0	175.97	182.96	6987.08	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	10	-0	175.97	185.83	7001.90	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	20	-0	175.97	188.70	7016.72	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	30	-0	175.97	191.57	7031.54	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	40	-0	175.97	194.44	7046.36	174.82	24.75	6197.83	1.146	-10.922	10.9610	3.786
1964	6	30	5	50	-0	175.97	197.31	7061.18	174.82	24.75	619				

(YEAR MO DY HR MN SEC		(PRINTING ANGLE COMPARISON - STATION AT ANDOVER ME)										(PRINTING ANGLE COMPARISON - STATION AT ANDOVER ME)											
(M-E-A-S-U-R-E-D		P-R-E-D-I-C-T-E-D					D-I-F-F-E-R-E-N-C-E-S					(M-E-A-S-U-R-E-D		P-R-E-D-I-C-T-E-D					D-I-F-F-E-R-E-N-C-E-S				
AZ-DEG		EL-DEG	RANGE-MI	AZ-DEG	EL-DEG	RANGE-MI	AZ-DIFF	EL-DIFF	RANGE-DIFF	ARC-DIF	RANGE-DIFF	AZ-DEG		EL-DEG	RANGE-MI	AZ-DIFF	EL-DIFF	RANGE-DIFF	ARC-DIF	RANGE-DIFF			
11964	6 30 5 10 -0	210.36	37.45	7446.59	210.36	37.45	7446.58	-0.000	0.001	0.001	0.0232	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
11964	6 30 5 20 -0	201.49	31.35	7347.55	201.68	31.32	7347.39	-0.012	0.012	0.012	0.0343	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012			
11964	6 30 5 30 -0	193.71	23.77	7214.50	193.69	23.71	7214.35	0.019	0.056	0.056	0.0581	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019			
11964	6 2 3 40 -0	275.88	25.01	6175.84	276.88	6.75	6314.68	-0.996	18.284	18.3091	-138.845	-0.996	18.284	18.3091	-138.845	-0.996	18.284	18.3091	-138.845	-0.996			
11964	6 2 3 42 -0	275.94	27.46	6186.80	277.00	10.00	6295.31	-1.060	17.962	17.962	-108.510	-1.060	17.962	17.962	-108.510	-1.060	17.962	17.962	-108.510	-1.060			
11964	6 10 7 52 -0	280.67	27.72	8201.62	277.11	13.21	8282.92	-1.141	17.617	17.6485	-80.800	-1.141	17.617	17.6485	-80.800	-1.141	17.617	17.6485	-80.800	-1.141			
11964	6 10 8 0 -0	274.59	30.17	8101.17	281.09	26.56	8308.63	-0.495	3.610	6.7579	-207.456	-0.495	3.610	6.7579	-207.456	-0.495	3.610	6.7579	-207.456	-0.495			
11964	6 10 8 10 -0	286.01	32.14	7859.83	273.45	29.37	8114.22	-7.444	2.771	6.9692	-254.388	-7.444	2.771	6.9692	-254.388	-7.444	2.771	6.9692	-254.388	-7.444			
11964	7 30 23 10 -0	287.17	15.74	4461.54	265.68	42.63	4483.39	21.493	-26.890	32.5607	-21.850	21.493	-26.890	32.5607	-21.850	21.493	-26.890	32.5607	-21.850	21.493			
11964	7 30 23 20 -0	270.42	37.43	4523.82	239.53	53.35	5007.62	30.895	-15.916	26.5686	-68.820	30.895	-15.916	26.5686	-68.820	30.895	-15.916	26.5686	-68.820	30.895			
11964	7 30 23 30 -0	246.88	49.81	4987.65	213.85	54.25	5676.42	33.029	-6.437	20.5984	-688.773	33.029	-6.437	20.5984	-688.773	33.029	-6.437	20.5984	-688.773	33.029			
11964	8 1 1 50 -0	272.56	17.52	6562.70	256.18	24.51	7019.15	16.375	-6.988	16.7892	-456.452	16.375	-6.988	16.7892	-456.452	16.375	-6.988	16.7892	-456.452	16.375			
11964	8 1 2 0 -0	260.38	21.80	7043.55	245.28	25.40	7515.27	15.100	-3.598	14.2879	-471.716	15.100	-3.598	14.2879	-471.716	15.100	-3.598	14.2879	-471.716	15.100			
11964	8 1 2 10 -0	249.56	23.23	7537.26	235.92	24.41	7971.31	13.639	-1.175	12.5275	-434.046	13.639	-1.175	12.5275	-434.046	13.639	-1.175	12.5275	-434.046	13.639			

ABOVE ELEMENTS COMPUTED FROM PRINTING ANGLES AND SLANT RANGES FROM ANDOVER ME LAT 44-63550 DEG WEST LONG 70-70030 DEG HT ABOVE MSL 945.00000 FT

Fig. 13 (cont.).

MEGEN G3 - ORBIT GENERATION FROM ANDOVER TRAJECTORIES - L C THOMAS *

RANGE DATA SUPPLIED ON ALL THREE POINTS. FREE FIT

REFRACTION CORRECTION APPLIED TO INPUT AND OUTPUT SATELLITE ELEVATIONS.

COMPUTED MODIFIED ORBITAL ELEMENTS-

PRINT ITER NO	EPOCH YEAR MO DY HR MINUTE	INCL DEGREES	ASC WEST LONG	NODE MINUTES	PSI DEG	LDAY- MIN	ARG PERIG DEG/TA	PERIGEE DEG	ANOM MIN	PERIOD MIN	PERIOD MIN/PERIOD	CHG ECCEN- TRICITY	RAD PERIGEE STAT MILES
1	4	1964 7 30 22 38.226	42.74888	218.36545	-8.15854	0.48840	0.19044	0.48840	225.26541	0.40101	0.40101	4565.221	
		1964 7 30 22 38.235	218.36768	218.36768	0.48840	0.19041	0.19041	0.19041	225.20241	FRM DATA POINT 1 TO 2	TIME		
		1964 7 30 22 38.239	218.36886	218.36886	0.48840	0.19044	0.19044	0.19044	225.16912	FRM DATA POINT 1 TO 3	TIME		

ORBIT OFFSET = 2.12031 DEG OF LAT. IF 0 1ST SUBSAT POINT IN ORBIT PLANE.

(YEAR MO DY HR MN SEC	(POINTING ANGLE COMPARISON - STATION AT ANDOVER ME)	D I F E R E N C E S)
(M-E-A-S-U-R-E-D	(P-R-E-D-I-C-T-E-D	(A-Z-DIFF EL-DIFF ARC-DIF RANGE-DIFF)
(AZ-DEG EL-DEG RANGE-MI	(AZ-DEG EL-DEG RANGE-MI	(AZ-DIFF EL-DIFF ARC-DIF RANGE-DIFF)
1964 7 30 23 10 -0	287.17 15.74 4461.54 287.17 15.74 4461.54	0.000 0.000 0.000
1964 7 30 23 20 -0	270.42 37.43 4523.82 270.41 37.44 4523.87	0.007 -0.006 0.0140
1964 7 30 23 30 -0	246.88 49.81 4987.65 246.86 49.80 4987.97	0.017 0.010 0.0185
1964 6 2 3 40 -0	275.88 25.01 6175.84 276.93 2.00 6362.61	-1.046 23.011 23.0332
1964 6 2 3 42 -0	275.94 27.96 6186.80 277.06 5.26 6333.38	-1.121 22.701 22.7258
1964 6 2 3 44 -0	275.97 30.83 6201.62 277.18 8.54 6311.29	-1.214 22.292 22.3210
1964 6 10 7 52 -0	280.67 27.72 8247.93 289.19 21.28 8483.61	-8.522 6.435 10.0705
1964 6 10 8 0 -0	274.59 30.17 8101.17 284.09 24.62 8394.27	-9.500 5.547 10.0888
1964 6 10 8 10 -0	266.01 32.14 7859.83 276.90 27.84 8226.51	-10.890 4.305 10.3611
1964 6 30 5 10 -0	210.36 37.45 7446.59 218.99 39.86 7524.03	-2.409 7.1512 77.435
1964 6 30 5 20 -0	201.69 31.35 7347.55 209.66 38.97 7452.25	-2.968 -3.618 7.5854
1964 6 30 5 30 -0	193.71 23.77 7214.50 201.16 28.59 7335.52	-7.454 -4.822 8.2423
1964 8 1 1 50 -0	272.56 17.52 6562.70 272.21 17.72 6571.79	-0.354 -0.195 0.3898
1964 8 1 2 0 -0	260.38 21.80 7043.55 260.05 21.87 7054.06	0.332 -0.071 0.3160
1964 8 1 2 10 -0	249.56 23.23 7537.26 249.25 23.24 7547.68	0.306 -0.008 0.2819

Fig. 14 - Fit of orbit generated from July 30 data only.

TELSTAR II - RATE GENERATION FROM DUAL MBE SETS L. C. THOMAS
COMPARISON FOR ALL THREE RANGE VALUES SUPPLIED.

RATES TO FOLLOW BASED ON FOLLOWING PAIR OF MODIFIED ORBITAL ELEMENTS-

REFERENCE TIME	1964 JUNE 30 2 HR 53-96210 MIN UT	1964 JULY 30 22 HR 38-22600 MIN UT
INCLINATION	42.76190 DEG	42.74888 DEG
ASCENDING NODE (WEST LONG)	219.33549 DEG	218.36545 DEG
PRIME SWEEP INTERVAL ONE DAY	-8.15271 MIN	-8.15854 MIN
PERIGEE AND SATELLITE POSITION	322.80277 DEG	0.48840 DEG
RATE OF CHANGE	0.19015 DEG/TA	0.19044 DEG/TA
ANOMALISTIC PERIOD	225.33698 MIN	225.26541 MIN
RATE OF CHANGE	0.	0.401010 MIN/TA
ECCENTRICITY	0.400788	0.401010
RADIUS OF PERIGEE	4567.91095 STATUTE MILES	4565.22095 STATUTE MILES

RATES CALCULATED FROM ABOVE ELEMENTS ARE

PRIME SWEEP INTERVAL ONE DAY -8.12511 MIN
 APSIDAL ADVANCE 0.19227 DEG/TA
 CHANGE IN ANOMALISTIC PERIOD 0.011360648 MIN/TA (PROVIDED ANOMALISTIC PERIOD OF INITIAL MBE SET IS CORRECT)
 ANOMALISTIC PERIOD OF INITIAL MBE SET = 225.33698 MIN

INCLINATION CHANGE -0.000066 DEG/TA
 ECCENTRICITY CHANGE 0.00000113 UNITS/TA
 RADIUS OF PERIGEE CHANGE -0.01372 STATUTE MI/TA

ABOVE RATES APPLIED TO FIRST SET OF MBES LISTED ABOVE WILL YIELD SECOND MBE SET.

FOLLOWING RATES APPLY IF GIVEN MBES CONNECTED WITH SINGLE ANOM PERIOD AND ZERO RATE OF CHANGE-

RATES CALCULATED FROM ABOVE ELEMENTS ARE

PRIME SWEEP INTERVAL ONE DAY -8.12511 MIN
 APSIDAL ADVANCE 0.19227 DEG/TA (0.1929 DEG/TA, SEE TEXT)
 CHANGE IN ANOMALISTIC PERIOD 0. MIN/TA
 ANOMALISTIC PERIOD = TIME BETWEEN MBE EPICHS/CAL NUMBER OF PERIGEE PASSES = 44384.263 / 196.00
 ANOMALISTIC PERIOD = 226.45032 MIN (225.300827 MIN, SEE TEXT)

INCLINATION CHANGE -0.000066 DEG/TA
 ECCENTRICITY CHANGE 0.00000113 UNITS/TA
 RADIUS OF PERIGEE CHANGE -0.01372 STATUTE MI/TA

ABOVE RATES APPLIED TO FIRST SET OF MBES LISTED ABOVE WILL YIELD SECOND MBE SET.

THEORETICAL VALUES FOLLOW-

REPURRIAN PERIOD= 225.38766 MIN
 NODAL PERIOD= 225.29030 MIN
 ANOMALISTIC PERIOD= 225.33692 MIN
 PRIME SWEEP INTERVAL= 1 DAY=8.15271 MIN
 APSIDAL ADVANCE= 0.19015 DEG/TA

ITERATIONS FOR PSI= 1
 ITERATIONS FOR APSIDAL ADVANCE= 2

GEN G3. The output is shown in Fig. 16. A considerable improvement in the fit over the June 2 to August 1 period is seen, with the bulk of the errors within the low hundredths of degree range. A maximum range error of only 7 miles over this interval is also noted.

Even these small errors may indicate the need for additional orbital rate improvement. The mean rates determined by combining June 2, 10, 30 and July 30 data are

$$\begin{aligned} \text{anomalistic period} &= 225.30049 \text{ minutes} \\ \text{apsidal advance rate} &= 0.19105^\circ \text{ per anomalistic period} \\ \text{prime sweep interval} &= \text{one day} - 8.12480 \text{ minutes.} \end{aligned}$$

Using these rates produces an even better trajectory fit, as shown in Fig. 17. The mere fact that the maximum measured error is about 0.05° and within five miles in range over a 3-month period indicates the near constancy of the secular rates and attests to the fact that little energy is being lost by the satellite as it orbits.

Fig. 18 illustrates the end result of the same type of procedure as described above, except that only one range point per determination is used in the generation. Here the pointing errors peak at 0.6° over the survey with maximum range errors about 25 miles.*

Fig. 19 repeats the above with absolutely no range data fed to the orbital generation process. Over the survey, maximum pointing and range errors are 0.8° and 76 miles respectively.

The advantage of 3 range points per determination over the several months' survey period is clearly evident by the results.

It must be stated that this illustrative example is somewhat simplified in that a certain amount of initial, free-fit orbit generation is omitted. In actual operation, the free-fit mode of MOEGEN G3 is often run for 6 to 12 data points in various combinations, and only the elements producing the best trajectory fit are preserved for later rate processing.

At this point, we intend to show how the techniques of Section VIII will provide a better round-the-world orbit fit by modifying the eccentricity and perigee radius of the orbital elements. Since no data pertaining to the TELSTAR satellites were readily obtainable from stations in southern latitudes, it was decided to proceed as described below.

A set of orbital elements† (which we shall call MO) was generated

* The -7449.740 and -7211.212 range "errors" indicated by Fig. 18 are not true errors but merely reflect the fact that no range data were supplied at these times for generation purposes. The true errors are ± 3 miles, as shown several lines below.

† These orbital elements had previously predicted Andover passes to a pointing accuracy of 0.07° .

MOEGEN G3 - ORBIT GENERATION FROM ANDOVER TRAJECTORIES - L C THOMAS *

3 PRINT RANGE DATA SUPPLIED. FIT FORCED TO JUNE 30 - JULY 30 RATES.

REFRACTION CORRECTION APPLIED TO INPUT AND OUTPUT SATELLITE ELEVATIONS.

COMPUTED MODIFIED ORBITAL ELEMENTS-

POINT NO	ITER NO	YEAR MO	DAY HR	MINUTE	INCL DEGREES	ASC WEST	NBDE LONG	PSI MINUTES	IDAY- ARG	PERIG DEG	PERIGEE CHG	ANBM PERIOD	PERIOD MIN	CHG DEG/DAY	ECCEN- TRICITY	RAD PERIGEE	STAT MILES
1	4	1964	6	30	2	53.984	42.76212	219.34173	-8.12511	322.80221	0.19129	225.30083	0.	0.40079	4567.873		
		1964	6	30	2	54.019	219.35058	322.80221	0.19124	225.24250	FRM DATA	POINT 1 TO 2	TIME				
		1964	6	30	2	54.011	219.34851	322.80221	0.19125	225.25618	FRM DATA	POINT 1 TO 3	TIME				

ORBIT OFFSET = 31.48778 DEG OF LAT. IF 0.1ST SUBSAT POINT IN ORBIT PLANE.

(YEAR MO DY HR MN SEC (POINTING ANGLE COMPARISON - STATION AT ANDOVER ME)

	M-E-A-S-U-R-E-D	P-R-E-D-I-C-T-E-D	D-I-F-F	F-E-R-R-E-S											
	AZ-DEG	EL-DEG	RANGE-MI	AZ-DEG	EL-DEG	RANGE-MI	AZ-DIFF	EL-DIFF	ARC-DIF	RANGE-DIFF					
(1964	6	30	5	10	-0	210.36	37.45	7446.59	210.36	37.43	7446.58	-0.000	0.021	0.0232	0.008
(1964	6	30	5	20	-0	201.69	31.35	7347.55	201.68	31.32	7347.43	0.009	0.031	0.0343	0.124
(1964	6	30	5	30	-0	193.71	23.77	7214.56	193.70	23.72	7214.44	0.013	0.052	0.0533	0.058
(1964	6	2	3	40	-0	275.88	25.01	6175.84	275.90	25.06	6182.44	-0.022	-0.045	0.0490	-0.598
(1964	6	2	3	42	-0	275.94	27.96	6184.80	275.96	28.00	6193.38	-0.031	-0.040	0.0442	-0.583
(1964	6	2	3	44	-0	275.97	30.83	6201.62	276.00	30.87	6208.18	-0.029	-0.038	0.0437	-0.564
(1964	6	10	7	52	-0	280.57	27.72	8247.93	280.69	27.73	8248.67	-0.017	-0.009	0.0198	-0.742
(1964	6	10	8	10	-0	274.59	30.17	8101.17	274.60	30.18	8101.09	-0.015	-0.009	0.0156	-0.749
(1964	6	10	8	10	-0	286.01	32.14	7859.83	286.01	32.13	7858.74	0.002	0.014	0.0198	1.093
(1964	7	30	25	10	-0	287.17	15.74	4461.54	287.25	15.65	4464.80	-0.080	0.087	0.1166	-3.258
(1964	7	30	25	20	-0	270.42	37.43	4523.82	270.53	37.37	4524.75	-0.114	0.062	0.1097	-0.932
(1964	7	30	25	30	-0	246.88	49.81	4987.65	247.01	49.78	4987.17	-0.135	0.035	0.0944	0.484
(1964	8	1	50	0	-0	272.56	17.52	6562.70	272.63	17.50	6562.56	-0.066	0.017	0.0660	0.136
(1964	8	1	2	0	-0	260.38	21.80	7043.55	260.44	21.75	7043.18	-0.060	0.047	0.0727	0.370
(1964	8	1	2	10	-0	249.56	23.23	7537.26	249.61	23.19	7536.95	-0.047	0.042	0.0602	0.313

ABOVE ELEMENTS COMPUTED FROM POINTING ANGLES
AND SLANT RANGES
FROM ANDOVER ME LAT 44.63550 DEG WEST LONG 70.70030 DEG HT ABOVE MSL 945.00000 FT

Fig. 16 - Fit obtained by combining July 30 and June 30 data.

NOEN G3 - ORBIT GENERATION FROM ANDBVER TRAJECTORIES - L C THOMAS *

RANGE DATA SUPPLIED ON ALL THREE PRINTS. FORCED FIT (AVERAGE RATES)

REFRACTION CORRECTION APPLIED TO INPUT AND OUTPUT SATELLITE ELEVATIONS.

COMPUTED MODIFIED ORBITAL ELEMENTS-

PRINT NO	ITER NO	E P O C H YEAR MO DY HR MINUTE	INCL DEGREES	ASC WEST LONG	PSI MINUTES	1DAY- DEG	ARG PERIG DEG	PERIG DEG	CHG ANOM MIN	PERIOD MIN	PERIOD MIN	CHG ECEN- MIN/PERIOD	TRICITY TRICITY	STAT MILES STAT MILES
1	4	1964 6 30 2 53.984	42.76210	219.36180	-8.12480	332.80243	0.19105	225.30049	0.	0.40079	4567.869			
		1964 6 30 2 54.020		219.35059	332.80243	0.19100		225.26189						
		1964 6 30 2 54.011		219.34859	332.80243	0.19101		225.25576						

ORBIT OFFSET = 31.48778 DEG BF LAT. IF 0 1ST SUBSAT PRINT IN ORBIT PLANE.

(YEAR MO DY HR MN SEC	(PRINTING ANGLE COMPARISON - STATION AT ANDBVER ME)	(M-E-A-S-U-R-E-D	(P-R-E-D-I-C-T-E-D	(D I F F	(F E R E N C E S)						
		AZ-DEG	EL-DEG	MI	AZ-DEG	EL-DEG	MI	AZ-DIFF	EL-DIFF	ARC-DIFF	RANGE-DIFF
11964 6 30 5 10 -0		210.36	37.45	7446.59	210.36	37.43	7446.58	-0.000	0.021	0.0221	0.008
11964 6 30 5 20 -0		201.69	31.35	7347.55	201.68	31.32	7347.43	0.009	0.031	0.0343	0.124
11964 6 30 5 30 -0		193.71	23.77	7214.50	193.70	23.72	7214.44	0.013	0.052	0.0537	0.058
11964 6 2 3 40 -0		275.88	25.01	6175.84	275.91	25.00	6180.25	-0.026	0.005	0.0242	-4.407
11964 6 2 3 42 -0		275.94	27.96	6186.80	275.97	27.95	6191.13	-0.026	0.008	0.0252	-4.327
11964 6 2 3 44 -0		275.97	30.83	6201.62	276.00	30.82	6205.87	-0.033	0.008	0.0305	-4.274
11964 6 10 7 52 -0		280.67	27.72	8247.93	280.69	27.73	8248.71	-0.020	-0.006	0.0185	-0.784
11964 6 10 8 0 -0		274.59	30.17	8101.17	274.61	30.18	8101.52	-0.017	-0.008	0.0185	-0.352
11964 6 10 8 10 -0		266.01	32.14	7859.83	266.02	32.13	7859.67	0.006	0.015	0.0198	0.162
11964 7 30 23 10 -0		287.17	15.74	4461.54	287.19	15.78	4465.22	-0.019	-0.042	0.0448	-3.676
11964 7 30 23 20 -0		270.42	37.43	4523.82	270.46	37.45	4527.84	-0.037	-0.024	0.0389	-4.023
11964 7 30 23 30 -0		246.88	49.81	4987.65	246.93	49.81	4991.24	-0.053	-0.004	0.0370	-3.592
11964 8 1 1 50 -0		272.56	17.52	6562.70	272.59	17.54	6566.02	-0.027	-0.018	0.0313	-3.324
11964 8 1 2 0 -0		260.38	21.80	7043.55	260.41	21.77	7046.47	-0.032	0.027	0.0408	-2.920
11964 8 1 2 10 -0		249.56	23.23	7537.26	249.59	23.20	7539.72	-0.028	0.029	0.0396	-2.461

ABOVE ELEMENTS COMPUTED FROM PRINTING ANGLES AND SLANT RANGES FROM ANDBVER ME LAT 44.63550 DEG WEST LONG 70.70030 DEG HT ABOVE MSL 945.00000 FT

Fig. 17 - Combined June 2, 10, 30 and July 30 data.

MORGEN G3 - ORBIT GENERATION FROM ANDOVER TRAJECTORIES - L C THOMAS *

RANGE DATA SUPPLIED ON CENTRAL PRINT ONLY. FORCED FIT.

REFRACTION CORRECTION APPLIED TO INPUT AND OUTPUT SATELLITE ELEVATIONS.

COMPUTED MODIFIED ORBITAL ELEMENTS-

PRINT NO	ITER NO	E P O C H YEAR MO DY HR MINUTE	INCL DEGREES	ASC WEST LONG	PSI MINUTES	IDAY- DEG	ARG PERIG	PERIGEE CHG ANOM	PERIOD MIN	PERIOD MIN/PERIOD	CHG TRICITY	ECCEN- STAT MILES	PERIOD 1 TO 2 TIME	PERIOD 2 TO 3 TIME
1	4	1964 6 30 2 53-749	42.72775	219.27264	-8.12480	322.62804	0.19105	225.30949	0.	0.40134	4566.226			
		1964 6 30 2 53-678	219.25497	322.62804	0.19115	225.31669	FROM DATA							
		1964 6 30 2 53-619	219.24006	322.62804	0.19123	225.31475	FROM DATA							

ORBIT OFFSET = 31.49270 DEG BE LAT. IF 0 1ST SUBSAT PRINT IN ORBIT PLANE.

(YEAR MO DY HR MN SEC	(PNTING ANGLE	COMPARISON	-	STATION AT	ANDOVER	ME)	J	F	E	R	E	N	C	E	S)
M-E-A-S-U-R-E-D	AZ-DEG	EL-DEG	RANGE-MI	AZ-DEG	EL-DEG	RANGE-MI	AZ-DEG	EL-DEG	EL-DIFF	EL-DIFF	ARC-DIFF	ARC-DIFF	RANGE-DIFF	RANGE-DIFF				
.11964 6 30 5 10 -0	210.36	37.45	0.	210.36	37.43	7449.74	-0.000	0.021	0.0221	-7449.740	0.0408	0.251						
.11964 6 30 5 20 -0	201.69	31.35	7347.55	201.67	31.31	7347.30	0.018	0.036	0.0791	-7211.212	0.0185	-11.663						
.11964 6 30 5 30 -0	193.71	23.77	0.	193.67	23.70	7211.21	0.039	0.070	0.0121	-9.237	0.0297	-5.853						
.11964 6 10 8 0 -0	280.67	27.72	8247.93	280.65	27.73	8259.59	0.016	-0.011	0.010	0.027	0.2783	-17.158						
.11964 6 10 8 10 -0	274.59	30.17	8101.17	274.58	30.17	8110.41	0.009	-0.003	0.041	-0.276	0.2619	-17.801						
.11964 6 2 3 40 -0	266.01	32.14	7859.83	266.00	32.11	7865.68	0.010	0.027	0.045	-0.259	0.2458	-18.371						
.11964 6 2 3 42 -0	275.88	25.01	6175.84	275.84	25.29	6193.00	0.041	-0.276	0.040	-0.243	0.0408	0.251						
.11964 6 2 3 44 -0	275.97	30.83	6186.80	275.90	28.22	6204.60	0.045	-0.259	0.000	0.021	0.0221	-3.150						
.11964 6 30 5 10 -0	210.36	37.45	7446.59	210.36	37.43	7449.74	-0.000	0.021	0.039	0.070	0.0408	0.251						
.11964 6 30 5 20 -0	201.69	31.35	7347.55	201.67	31.31	7347.30	0.018	0.036	0.039	0.070	0.0408	0.251						
.11964 6 30 5 30 -0	193.71	23.77	7214.50	193.67	23.70	7211.21	0.039	0.070	0.039	0.070	0.0408	0.251						
.11964 7 30 23 10 -0	287.17	15.74	4461.54	286.87	16.32	4467.66	0.370	-0.387	0.297	-0.584	0.6496	-6.121						
.11964 7 30 23 20 -0	270.42	37.43	4523.82	270.05	37.82	4543.36	0.370	-0.387	0.378	-0.387	0.4852	-19.539						
.11964 7 30 23 30 -0	246.88	49.81	4987.65	246.50	49.97	5012.97	0.176	-0.159	0.176	-0.159	0.2311	-25.348						
.11964 8 1 1 50 -0	272.56	17.52	6562.70	272.38	17.68	6585.20	0.119	-0.058	0.119	-0.058	0.1249	-22.648						
.11964 8 1 2 0 -0	260.38	21.80	7043.55	260.26	21.86	7066.20	0.075	-0.023	0.075	-0.023	0.0730	-21.223						
.11964 8 1 2 10 -0	249.56	23.23	7537.26	249.49	23.25	7558.48	0.075	-0.023	0.075	-0.023	0.0730	-21.223						

ABOVE ELEMENTS COMPUTED FROM PRINTING ANGLES
FROM ANDOVER ME LAT 44.63550 DEG WEST LONG 70.70030 DEG HT ABOVE MSL 945.00000 FT

Fig. 18 - Fit obtained with only one range point per determination: June 2, 10, 30 and July 30 data.

MOEGEN G3 - BRBIT GENERATION FROM ANDOVER TRAJECTORIES - L C THOMAS *

NO RANGE DATA SUPPLIED TO PROGRAM FOR BRBIT GENERATION FORCED FIT.

REFRACTION CORRECTION APPLIED TO INPUT AND OUTPUT SATELLITE ELEVATIONS.

COMPUTED MODIFIED ORBITAL ELEMENTS-

POINT ITER E P C H INCL ASC NBDP PSI LDAY- ARG PERIG PERIGEE CHG ANOM PERIØD PERIØD CHG ECCEN- RAD PERIGEE
 NO YEAR MO DY HR MINUTE DEGREES WEST LONG MINUTES DEG DEG/TA MIN MIN/PERIØD FRICITY STAT MILES
 1 4 1964 6 30 2 53.023 42.68924 218.92101 -8.12480 321.70110 0.19105 225.30049 0. 0.39594 4608.807
 1964 6 30 2 50.823 321.31743 321.70110 0.19440 225.24918 FROM DATA POINT 1 TO 2 TIME
 1964 6 30 2 50.597 218.31086 321.70110 0.19443 225.29213 FROM DATA POINT 1 TO 3 TIME
 BRBIT OFFSET = 31.39869 DEG OF LAT. IF 0 1ST SUBSAT POINT IN ORBIT PLANE.

(YEAR MO DY HR MN SEC	(POINTING ANGLE COMPARISON - STATION AT ANDOVER ME)	D I F	F E R E N C E S)
AZ-DEC EL-DEC RANGE-MI	P-R-E-D-I-C-T-E-D	EL-DIFF	ARC-DIF RANGE-DIFF
210.36 37.45 0. 210.36 37.43 7389.70	-0.000	0.021	0.0221-285.073
1964 6 30 5 10 -0	210.36 37.45 0. 210.36 37.43 7389.70	-0.000	0.021
1964 6 30 5 20 -0	201.69 31.35 0. 201.47 31.21 7284.01	0.224	0.138
1964 6 30 5 30 -0	193.71 23.77 0. 193.29 23.44 7149.13	0.423	0.329
1964 6 2 3 40 -0	275.88 25.01 6175.84 275.75 24.79 6246.59	0.134	0.218
1964 6 2 3 42 -0	275.94 27.96 6186.80 275.81 27.69 6255.35	0.133	0.271
1964 6 2 3 44 -0	275.97 36.83 6201.62 275.85 30.51 6267.71	0.123	0.318
1964 6 10 7 52 -0	280.67 27.72 8247.93 280.97 27.46 8233.58	-0.305	0.257
1964 6 10 8 0 -0	274.59 30.17 8101.17 274.86 29.97 8073.53	-0.273	0.199
1964 6 10 8 10 -0	266.01 32.14 7859.83 266.18 31.99 7816.45	-0.175	0.152
1964 6 10 8 10 -0	210.36 37.45 7446.59 210.36 37.43 7389.70	0.000	0.021
1964 6 30 5 20 -0	201.69 31.35 7347.55 201.47 31.21 7284.01	0.224	0.138
1964 6 30 5 30 -0	193.71 23.77 7214.50 193.29 23.44 7149.13	0.423	0.329
1964 7 30 23 10 -0	287.17 15.74 4661.54 286.97 15.58 4630.10	0.202	0.840
1964 7 30 23 20 -0	270.42 37.43 4523.82 270.62 37.47 4599.69	-0.195	0.245
1964 7 30 23 30 -0	246.88 49.81 4987.65 247.63 49.86 5048.65	-0.750	0.406
1964 8 1 1 50 -0	272.56 17.52 5062.70 272.99 17.61 6618.74	-0.434	0.089
1964 8 1 2 0 -0	240.38 21.80 7043.55 240.94 21.83 7081.45	-0.560	0.034
1964 8 1 2 10 -0	249.36 23.23 7537.26 250.15 23.30 7556.45	-0.586	0.067

ABOVE ELEMENTS COMPUTED FROM POINTING ANGLES FROM ANDOVER ME LAT 44.63550 DEG WEST LONG 70.70030 DEG HT ABOVE MSL 945.00000 FT

Fig. 19 - Same as Fig. 18, but with no range data.

TABLE II — ALTERATIONS TO MO ELEMENTS

Data Point of 4th Trajectory	Azimuth Alteration	Elevation Alteration	Range Alteration
1	0.21°	-0.07°	4.85 miles
2	0.06°	-0.02°	1.03 miles
3	-0.10°	+0.02°	-0.87 mile

from Andover data and used to simulate trajectories over Andover for a sequence of four passes. The computed pointing angles for these passes then were rounded off to the nearest hundredth of a degree so as to resemble typically reported data from Andover. Three points from each of the first 3 passes were then selected for submission to the MOEGEN G3 program. Three points from the fourth pass were purposely altered before submission to simulate the effects of tracking errors, mechanical jitter of the ground antenna, errors in assessment of refraction which affect measured position, and digital noise in the equipment at the ground station. These alterations were slight and are given in Table II. Using the data from the 4 passes, MOEGEN G3 produced 4 sets of orbital elements.

Fig. 20 displays the generated elements M1, M2, M3 and M4. For each of these elements, the first 3 lines of data under "pointing angle comparisons" (see Fig. 20) show the differences between the input data points of the submitted trajectories and those predicted by the derived elements. For each set of elements, the pointing error is within 0.07° and the range difference under 0.9 miles. Looking at each of the *first 3* comparison lines, all 4 sets of elements produce excellent fits to reported Andover trajectories, and from this examination there would be no reason to expect that any particular set of elements would be decidedly better than any other. However, comparisons of the M4 orbit with 6/30 trajectory data indicate that something is wrong with these elements because of the increase in pointing angle differences on this date. This, however, could be produced by variety of causes: e.g., errors in epoch, rates, ellipse shape, size and orientation of the orbit. The point to make is that the M4 orbit fits well the trajectory over Andover from which it was derived; the purposely introduced errors in that trajectory have not done violence to the orbit theory, so that the motion of the satellite is in essential accord with those motions permitted by the earth's gravitational potential.

Using the same MO elements as previously, trajectories over Johannes-

MDEGEN G3 - BRBIT GENERATION FROM ANDOVER TRAJECTORIES - L C THOMAS *

3 PRINT RANGE DATA. FIT FORCED TO AVERAGE RATES.

REFRACTION CORRECTION APPLIED TO INPUT AND OUTPUT SATELLITE ELEVATIONS.

COMPUTED MODIFIED ORBITAL ELEMENTS-

POINT ITER E P O C H INCL ASC NODE PSI LDAY- ARG PERIG PERIG CHG ANUM PERIØD PERIØD CHG ECCEN- RAD PERIGEE
 NO YEAR MO DY HR MINUTE DEGREES WEST LONG MINUTES DEG DEG/TA MIN MIN PERIØD TRICITY STAT MILES
 1 4 1964 6 30 17 55-214 42-74526 85-92082 -8-12480 323-58760 0.19105 225-30049 0. 0-40099 4564.555 -MI
 1964 6 30 17 55-224 85-92345 323-58760 0.19094 225-17527 FROM DATA PRINT 1 TO 2 TIME
 1964 6 30 17 55-228 85-92430 323-58760 0.19091 225-13461 FROM DATA PRINT 1 TO 3 TIME
 BRBIT ØFFSET = 24-26940 DEG ØF LAT. IF Ø 1ST SUBSAT PRINT IN ØRBIT PLANE.

(YEAR MO DY HR MN SEC	(ME-A-S-U-R-E-D	P-R-E-O-I-C-T-E-D	D I F F E	R E N C E S
		AZ-DEG EL-DEG RANGE-MI	AZ-DEG EL-DEG RANGE-MI	AZ-DIFF EL-DIFF	ARC-DIF RANGE-DIFF
(1964 6 30 18 14 -0		172.10 25.01 2526.48	172.10 24.98 2526.48	-0.000 0.035	0.0363 0.004
(1964 6 30 18 24 -0		113.28 41.45 3089.51	113.29 41.43 3089.32	-0.007 0.017	0.0221 0.186
(1964 6 30 18 34 -0		82.95 32.12 4435.73	82.96 32.10 4435.08	-0.005 0.019	0.0221 0.652
(1964 6 30 18 14 -0		172.10 25.01 2526.48	172.10 24.98 2526.48	-0.000 0.035	0.0363 0.004
(1964 6 30 18 24 -0		113.28 41.45 3089.51	113.29 41.43 3089.32	-0.007 0.017	0.0221 0.186
(1964 6 30 18 34 -0		82.95 32.12 4435.73	82.96 32.10 4435.08	-0.005 0.019	0.0221 0.652
(1964 6 30 22 20 -0		253.31 67.71 3592.57	253.32 67.69 3592.28	-0.005 0.024	0.0262 0.291
(1964 6 30 22 30 -0		134.39 85.02 4284.55	134.52 85.02 4283.77	-0.129 -0.005	0.0140 0.775
(1964 6 30 22 40 -0		96.87 69.49 5133.89	96.88 69.49 5132.73	-0.010 -0.005	0.0140 1.162
(1964 7 1 2 30 -0		275.16 46.86 5950.66	275.16 46.85 5949.91	-0.003 0.009	0.0140 0.754
(1964 7 1 2 40 -0		262.01 56.52 6228.15	262.02 56.49 6227.08	-0.014 0.027	0.0297 1.071
(1964 7 1 2 50 -0		243.63 62.12 6516.14	243.65 62.10 6514.79	-0.023 0.020	0.0252 1.348
(1964 7 1 6 50 -0		278.01 14.41 8863.39	278.01 14.40 8862.02	-0.004 0.007	0.0121 1.367
(1964 7 1 7 0 -0		270.85 15.34 8919.85	270.85 15.34 8918.32	-0.003 0.004	0.0121 1.530
(1964 7 1 7 10 -0		263.66 15.23 8909.19	263.66 15.16 8907.54	-0.004 0.065	0.0660 1.647

Fig. 20 - Modified orbital elements generated from simulated Andover passages.

POINT NB	ITER NB	E P O C H YEAR MB DY HR MN SEC	INCL DEGREES	ASC DEG	NODE DEG	PSI DEG/TA	IDAY- MIN	ARG PERIOD	PERIGEE CHG ANOM	PERIOD MIN	PERIGEE CHG ECCEN-	RAD PERIGEE STAT MILES	TRICITY MIN/PERIOD	M2
1	4	1964 6 30 21 40-503	42.74677	-8.12480	323.75162	0.19105	225.30049	0.	0.40110	4563.160				
		1964 6 30 21 40-543	142.5847	323.75162	0.19086	225.07273	FRM DATA	POINT 1 TO 2 TIME						
		1964 6 30 21 40-552	142.57085	323.75162	0.19081	225.01965	FRM DATA	POINT 1 TO 3 TIME						
		RRBIT OFFSET = 3.77641 DEG OF LAT. IF 0 1ST SUBSAT POINT IN ORBIT PLANE.												
		(YEAR MB DY HR MN SEC	(PRINTING ANGLE COMPARISON - STATION AT ANDOVER ME)											
		(ME-A-S-U-R-E-D	(P-R-E-D-I-C-T-E-D											
		AZ-DEG	EL-DEG	RANGE-MI	AZ-DEG	EL-DEG	RANGE-MI	AZ-DIFF	EL-DIFF	ARC-DIF	RANGE-DIFF	F	E	R
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.021			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		113.28	41.45	3089.51	113.25	41.42	3090.01	0.032	0.026	0.0370	-0.504			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0280	-0.164			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.015	0.0490	0.703			
		82.95	32.12	4435.73	82.94	32.09	4435.89	0.006	0.026	0.0370	-0.504			
		253.31	67.71	3592.57	253.31	67.70	3592.56	0.000	0.007	0.0140	0.010			
		134.39	65.02	4284.55	134.46	85.02	4284.26	-0.070	-0.004	0.0156	0.291			
		96.87	69.49	5133.89	96.87	69.50	5133.07	-0.002	-0.009	0.0156	0.822			
		172.10	25.01	2526.48	172.05	24.99	2525.78	0.051	0.0					

POINT NB	ITER NB	YEAR	MO	DAY	HR	MIN	SEC	INCL DEGREES	ASC WEST	NODE LONG	PSI MINUTES	IOAY- MINUTES	ARG DEG	PERIG DEG	PERIGEE DEG/TA	CHG MIN	ANOM MIN	PERIOD MIN	PERIOD TRICITY	CHG MIN	ECCEN- TRICITY	RAD PERICEE STAT MILES	
1	4	1964	7	1	5	11	133	42.72288	255.82689	-8.12480	324.12966	0.19105	225.30049	D.	0.41158	4450.926							
		1964	7	1	5	11	481	324.12966	0.19038														
		1964	7	1	5	11	473	324.12966	0.19039														
		ORBIT OFFSET = 15.26678 DEG OF LAT. IF 0 1ST SUBSAT POINT IN ORBIT PLANE.																					
		PRINTING ANGLE COMPARISON - STATION AT ANDOVER ME)																					
		(YEAR	MO	DAY	HR	MIN	SEC	(ME- AZ-DEG	A-SU-R-E-D	P-R-E-D-I-C-T-E-D	D	I	F	F	E	R	E	N	C	E	S	
		(1964	7	1	6	50	-0	277.80	14.48	8858.34	277.80	14.48	8858.54	-0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		(1964	7	1	7	0	-0	270.79	15.36	8918.82	270.81	15.35	8919.04	-0.021	0.006	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185
		(1964	6	30	18	14	-0	263.76	15.21	8910.06	263.80	15.15	8910.24	-0.043	0.056	0.0699	0.0699	0.0699	0.0699	0.0699	0.0699	0.0699	0.0699
		(1964	6	30	18	24	-0	113.28	41.45	3089.51	109.84	39.88	3081.05	3.438	1.566	3.0420	3.0420	3.0420	3.0420	3.0420	3.0420	3.0420	3.0420
		(1964	6	30	18	34	-0	82.95	32.12	4435.73	81.56	30.45	4469.52	1.390	1.566	2.0475	2.0475	2.0475	2.0475	2.0475	2.0475	2.0475	2.0475
		(1964	6	30	22	20	-0	253.31	67.71	3592.57	251.68	70.04	3535.27	1.635	2.328	2.4012	2.4012	2.4012	2.4012	2.4012	2.4012	2.4012	2.4012
		(1964	6	30	22	30	-0	134.39	85.02	4284.55	121.53	83.50	4272.76	12.860	1.522	1.9842	1.9842	1.9842	1.9842	1.9842	1.9842	1.9842	1.9842
		(1964	6	30	22	40	-0	96.87	69.49	5133.89	96.34	68.05	5144.37	0.535	1.444	1.4571	1.4571	1.4571	1.4571	1.4571	1.4571	1.4571	1.4571
		(1964	7	1	2	30	-0	275.16	46.86	5950.66	274.19	47.76	5913.86	0.965	-0.904	1.1151	1.1151	1.1151	1.1151	1.1151	1.1151	1.1151	1.1151
		(1964	7	1	2	40	-0	262.01	56.52	6228.15	260.79	57.08	6209.12	1.216	0.564	0.8727	0.8727	0.8727	0.8727	0.8727	0.8727	0.8727	0.8727
		(1964	7	1	2	50	-0	243.63	62.12	6516.14	242.38	62.35	6508.58	1.255	-0.233	0.6293	0.6293	0.6293	0.6293	0.6293	0.6293	0.6293	0.6293
		(1964	7	1	6	50	-0	278.01	14.41	8863.39	277.80	14.48	8858.54	0.210	0.270	0.2149	0.2149	0.2149	0.2149	0.2149	0.2149	0.2149	0.2149
		(1964	7	1	7	0	-0	270.85	15.34	8919.85	270.81	15.35	8919.04	0.039	-0.014	0.0402	0.0402	0.0402	0.0402	0.0402	0.0402	0.0402	0.0402
		(1964	7	1	7	10	-0	263.66	15.23	8909.19	263.80	15.15	8910.24	-0.143	0.076	0.1581	0.1581	0.1581	0.1581	0.1581	0.1581	0.1581	0.1581

Fig. 20 (cont.).

burg, South Africa were produced. These were compared with the predictions from the generated M4 elements in a computer comparison program. The results are given in Fig. 21, where all listed data are for satellite positions not visible to Andover (subsatellite latitudes 4° to -42°). The average pointing error is 2.7° and the root-mean-square range error is 65.5 miles.

The four generated elements were then submitted to the MOEFIT program, which applies the principles of Section VIII to generate an ellipse of best fit using only those portions of the orbits, described by the input elements, which are visible from Andover. Fig. 22 shows the eccentricity of 0.400593 and the perigee radius of 4565.168 statute miles determined from this ellipse. It also indicates that the root-mean-square departure of the fitted ellipse from those points generated by the four elements is in the order of 10^{-7} . Fig. 23 graphs the fitted ellipse as a curved line, while the points visible from Andover and predicted by the individual elements are noted by zeros on the plot.

The M4 set of elements was then altered by substituting the eccentricity and radius of perigee values from the MOEFIT program and was again compared with Johannesburg trajectories given by the original set of elements, MO. Typical results are given in Fig. 24, where it is seen that the average pointing angle error is now 0.07° , about 2 orders

ROUND-THE-WORLD-ORBIT-TEST - MOES GENERATED FROM ANDOVER L C THOMAS

ORBITAL ELEMENTS M-4 FROM MOEGEN G3 USED.

STATION AT JOBBURG, S.A.F LAT= -26.03246 DEG WEST LONG = 331.75878 DEG HT ABOVE MSL= 5370.00000 FT															
YEAR	MO	DAY	HR	MIN	SEC	SAT P8S FROM MBE 1			SAT P8S FROM DATA			D I F F E R E N C E S			
						AZ-DEG	EL-DEG	RANGE-MI	AZ-DEG	EL-DEG	RANGE-MI	AZ-DIFF	EL-DIFF	ARC-DIF	RANGE-DIFF
1964	6	30	1	50	0	288.75	11.10	7605.79	288.53	11.98	7559.01	-0.221	0.878	0.9043	-46.777
1964	6	30	2	0	0	284.15	18.01	6463.13	283.78	19.20	6411.44	-0.367	1.191	1.2407	-51.693
1964	6	30	2	10	0	277.26	26.86	5135.54	276.55	28.50	5085.55	-0.713	1.635	1.7529	-49.993
1964	6	30	2	20	0	263.52	39.54	3647.60	261.69	41.83	3616.87	-1.829	2.287	2.6741	-30.732
1964	6	30	2	30	0	213.29	54.33	2253.39	206.47	55.47	2295.77	-6.816	1.137	4.0786	42.382
1964	6	30	2	40	0	136.40	14.84	2296.01	134.91	14.78	2430.20	-1.487	-0.057	1.4384	134.186
1964	7	1	4	40	0	251.49	2.50	5127.70	250.75	4.33	5068.46	-0.737	1.834	1.9758	-59.242
1964	7	1	4	50	0	236.65	14.72	3173.39	235.10	17.25	3136.97	-1.549	2.525	2.9315	-36.423
1964	7	1	5	0	0	174.13	27.48	1518.65	169.84	29.93	1586.01	-4.293	2.448	4.4906	67.362
1964	7	1	8	56	0	277.23	31.34	832.02	277.78	37.04	906.03	0.549	5.697	5.7154	74.006

MODIFIED ORBITAL ELEMENTS FOR ABOVE TABLE ARE-

AVG = 2.7202 65.541 = RMS

ELEMENTS NO 1
 REFERENCE TIME 1964 7 1 5 HR 11.13300 MIN UT
 INCLINATION 42.72288 DEG
 ASCENDING NODE WEST LONG 255.82689 DEG
 PRIME SWEEP INTERVAL = 1 DAY -8.124800 MIN
 PERIGEE AND SATELLITE ARGUMENT 324.12966 DEG
 RATE OF CHANGE 0.191050 DEG PER PERIOD
 ANOMALISTIC PERIOD 225.30049 MIN
 RATE OF CHANGE. MIN PER PERIOD
 ECCENTRICITY 0.4115810
 RADIUS OF PERIGEE 4450.926 STATUTE MILES

Fig. 21—Comparison of M4 elements to Johannesburg trajectories derived from MO.

ROUND-THE-WORLD-ORBIT-TEST - MOES GENERATED FM ANDOVER L C THOMAS
 SUCCESSION OF PASSES ON 6/30 AND 7/1/1964 USED FOR R.P.E DETERMINATION
 ABS VALUE OF MAX ERROR= 5.8822479E-07
 LEAST SQUARES ORBITAL ELEMENT VALUES- MEAN SQUARE ERROR= 2.1384752E-14
 RMS ERROR= 1.4623526E-07
 ECCENTRICITY = 0.400593 RADIUS OF PERIGEE = 4565.1683 STATUTE MILES
 SEMI-MAJOR AXIS = 7616.1472 STATUTE MILES

AVERAGE PERIGEE POSITION = 324.14983 DEG AT EPOCH OF FIRST MOES

INPUT MODIFIED ORBITAL LIST FOLLOWS-

REFERENCE TIME 1964 6 30 17 HOUR 55.21400 MIN U.T.
 INCLINATION 42.74526 DEG
 ASCENDING NODE LONG 85.92082DEG WEST
 PRIME SWEEP INTERVAL ONE DAY -8.12480 MIN
 PERIGEE AND SATELLITE POSITION 323.58760 DEG
 RATE OF CHANGE 0.19105 DEG PER PERIOD
 ANOMALISTIC PERIOD 225.30049 MIN
 RATE OF CHANGE 0. MIN PER PERIOD
 ECCENTRICITY 0.40099
 RADIUS OF PERIGEE 4564.55493 STATUTE MILES

M1

START CARD IMAGE IS-
 1964 6 30 18 7 0 70 60 60

REFERENCE TIME 1964 6 30 21 HOUR 40.50300 MIN U.T.
 INCLINATION 42.74677 DEG
 ASCENDING NODE LONG 142.55847DEG WEST
 PRIME SWEEP INTERVAL ONE DAY -8.12480 MIN
 PERIGEE AND SATELLITE POSITION 323.75162 DEG
 RATE OF CHANGE 0.19105 DEG PER PERIOD
 ANOMALISTIC PERIOD 225.30049 MIN
 RATE OF CHANGE 0. MIN PER PERIOD
 ECCENTRICITY 0.40110
 RADIUS OF PERIGEE 4563.15997 STATUTE MILES

M2

START CARD IMAGE IS-
 1964 6 30 22 1 0 124 120 120

REFERENCE TIME 1964 7 1 1 HOUR 25.83200 MIN U.T.
 INCLINATION 42.74387 DEG
 ASCENDING NODE LONG 199.19972DEG WEST
 PRIME SWEEP INTERVAL ONE DAY -8.12480 MIN
 PERIGEE AND SATELLITE POSITION 323.98413 DEG
 RATE OF CHANGE 0.19105 DEG PER PERIOD
 ANOMALISTIC PERIOD 225.30049 MIN
 RATE OF CHANGE 0. MIN PER PERIOD
 ECCENTRICITY 0.40093
 RADIUS OF PERIGEE 4566.15997 STATUTE MILES

M3

START CARD IMAGE IS-
 1964 7 1 2 2 0 138 120 120

REFERENCE TIME 1964 7 1 5 HOUR 11.13300 MIN U.T.
 INCLINATION 42.72288 DEG
 ASCENDING NODE LONG 255.82689DEG WEST
 PRIME SWEEP INTERVAL ONE DAY -8.12480 MIN
 PERIGEE AND SATELLITE POSITION 324.12966 DEG
 RATE OF CHANGE 0.19105 DEG PER PERIOD
 ANOMALISTIC PERIOD 225.30049 MIN
 RATE OF CHANGE 0. MIN PER PERIOD
 ECCENTRICITY 0.41158
 RADIUS OF PERIGEE 4450.92596 STATUTE MILES

M4

START CARD IMAGE IS-
 1964 7 1 6 14 0 99 60 60

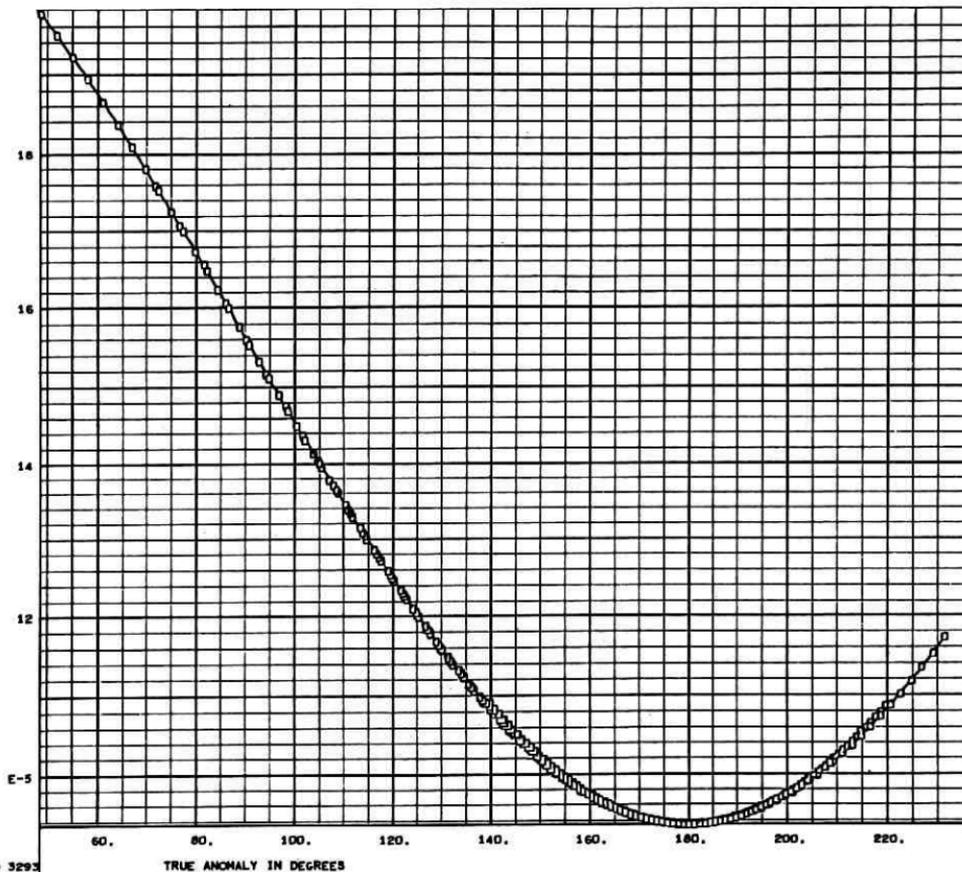
Fig. 22 — Eccentricity and perigee radius from ellipse of least-squares fit.

of magnitude better than the value before ellipse fitting. The root-mean-square range error has decreased an order of magnitude and is now 5.2 statute miles.

It is to be noted that only four passes at Andover were used in this illustrative procedure, since these four encompassed the entire portion of the orbit visible from Andover. However, as apsidal advance takes place,

new portions of the orbit become visible and, provided little orbit decay occurs, these may also be used to obtain even better ellipse fits and further refine values for the eccentricity and perigee radius.

It is also noted that in actual practice each of the trajectories used would have suffered slight measurement errors akin to those artificially introduced into the fourth trajectory that produced elements M4. Nonetheless, the derived elements would predict well that portion of the orbit ellipse from which they came (as was demonstrated by M4) and would therefore be improved by MOEFIT for round-the-world use by about the same amounts as shown in the illustrated example.



JOB NO 3293
09HRS
55MIN
56SEC
10/27/64
PAGE NO 1

ROUND-THE-WORLD-ORBIT-TEST - MOES GENERATED FM ANDOVER L C THOMAS
SUCCESSION OF PASSES ON 6/30 AND 7/1/1964 USED FOR RP,E DETERMINATION

Fig. 23 — Ellipse of least-squares fit to elements.

XII. CONCLUSIONS

The methods described have been used extensively for those orbit determinations of the TELSTAR satellites dealing with the satellite communications experiment. The following merits of this orbital technique are noted:

(i) The elements chosen explicitly express the secular rates, thereby simplifying and making more economical the generation of predicted trajectories and drive tapes for the Andover antenna.

(ii) The generation procedure is geared to sparse and discontinuous tracking data.

(iii) Using a modified Gaussian technique for orbit generation provides a straightforward method for range determination to which orbital perturbations are easily added on an iterative basis.

(iv) The computation of secular rates only saves time and money and proves adequate for the tracking requirements of the TELSTAR satellite experiments (0.05° pointing accuracy).

(v) For cases where the secular rates are not constant over a given span of time, the methods described will still function provided tracking data are available frequently enough to assure essentially constant rates from each data acquisition time to the next.

(vi) The method provides for the operational inclusion of perturbations

AROUND-THE-WORLD-ORBIT-TEST - MBS GENERATED FM ANDOVER L C THOMAS

M-4 ELEMENTS WITH ECCENTRICITY AND RADIUS OF PERIGEE FROM MBEFIT.

STATION AT JOBBURG, S. AF LAT= -26.03246 DEG WEST LONG = 331.75878 DEG HT ABOVE MSL= 5370.00000 FT

YEAR	MO	DY	HR	MN	SEC	SAT POS FROM MBE 1			SAT POS FROM DATA			DIFFERENCES			
						AZ-DEG	EL-DEG	RANGE-MI	AZ-DEG	EL-DEG	RANGE-MI	AZ-DIFF	EL-DIFF	ARC-DIF	RANGE-DIFF
1964	6	30	1	50	0	288.56	11.97	7550.87	288.53	11.98	7559.01	-0.027	0.009	0.0280	8.142
1964	6	30	2	0	0	283.81	19.20	6403.74	283.78	19.20	6411.44	-0.027	0.003	0.0242	7.704
1964	6	30	2	10	0	276.59	28.51	5078.57	276.55	28.50	5085.55	-0.040	-0.007	0.0370	6.983
1964	6	30	2	20	0	261.72	41.85	3611.05	261.69	41.83	3616.87	-0.034	-0.017	0.0321	5.821
1964	6	30	2	30	0	206.38	55.52	2292.20	206.47	55.47	2295.77	0.091	-0.047	0.0699	3.566
1964	6	30	2	40	0	134.85	14.72	2429.98	134.91	14.78	2430.20	0.055	0.064	0.0833	0.222
1964	7	1	4	40	0	250.78	4.16	5062.97	250.75	4.33	5068.46	-0.029	0.168	0.1700	5.492
1964	7	1	4	50	0	235.13	17.22	3133.32	235.10	17.25	3136.97	-0.026	0.026	0.0350	3.653
1964	7	1	5	0	0	169.84	29.94	1584.84	169.84	29.93	1586.01	0.000	-0.012	0.0171	1.166
1964	7	1	8	56	0	277.52	36.95	908.26	277.78	37.04	906.03	0.264	0.093	0.2305	-2.234

MODIFIED ORBITAL ELEMENTS FOR ABOVE TABLE ARE-

ELEMENTS NO	1
REFERENCE TIME	1964 7 1 5 HR 11.13300 MIN UT
INCLINATION	42.72288 DEG
ASCENDING NODE WEST LONG	255.82689 DEG
PRIME SWEEP INTERVAL	= 1 DAY -8.124800 MIN
PERIGEE AND SATELLITE ARGUMENT	324.12966 DEG
RATE OF CHANGE	0.191050 DEG PER PERIOD
ANOMALISTIC PERIOD	225.30049 MIN
RATE OF CHANGED.	MIN PER PERIOD
ECCENTRICITY	0.4005930
RADIUS OF PERIGEE	4565.168 STATUTE MILES

AVG=0.0727 5.209=RMS

Fig. 24 — Comparison of M4 elements (with fitted ellipse corrections) to Johannesburg trajectories.

not computed here. For example, cyclic perturbations may be expressed, if indicated in the data, by generation of a multiplicity of elements over a time span and noting the derivatives of the "secular" rates.

XIII. ACKNOWLEDGMENTS

The author gratefully acknowledges his indebtedness to all those authors listed in the references, particularly those greats of the past — such as Gauss and LaPlace — who, under less favorable conditions, produced the earlier "breakthroughs" from which the lesser of us now derive benefit.

APPENDIX A

Relationship of the Triangular Area between Two Radius Vectors and the Orbit Normal

Consider triangle 023 of Fig. 4 and project it upon the yz plane. The vertices of the projected triangle are, therefore, $(0,0,0)$, $(0,y_3,z_3)$, $(0,y_2,z_2)$ and hence its area becomes

$$A_p = \frac{1}{2}(y_2z_3 - z_2y_3). \quad (119)^*$$

Now the normal to the yz plane is, of course, the x axis. The normal to the orbit plane, shown in Fig. 4, makes an angle n with the x axis. Since the area of triangle 023 is the projected area referred to above divided by the cosine of the angle between the orbit plane and the yz plane, and since this angle is the same as the angle n between the normals to the two planes, it follows directly that the area of 023 is

$$A_{023} = \frac{1}{2}(y_2z_3 - z_2y_3)/\cos n \quad (120)$$

or

$$y_2z_3 - z_2y_3 = A_{23} \cos n \quad (9)$$

where

$$A_{23} = 2A_{023}. \quad (121)$$

APPENDIX B

Relations Involving the Earth's Oblateness

Because the earth is not a perfect sphere, two corrections are normally applied to satellite positional data as measured from any specified ground

* See page 266, Ref. 23.

station. The same corrections in the inverse sense are applied when trajectory predictions are made from orbital elements. The first of these recognizes that the geocentric distance of the station is a function of geodetic latitude; the second correction relates to the change in azimuth and elevation of the satellite when referred to a spherical as opposed to an oblate earth.

The geocentric distance⁵ of the station is simply

$$R = H + R_E(0.998320047 + 0.001683494 \cos 2\varphi - 0.000003549 \cos 4\varphi + 0.000000008 \cos 6\varphi) \quad (125)$$

where

H = the height of the station above mean sea level in statute miles

R_E = the earth's equatorial radius = 3963.347 statute miles

φ = the geodetic latitude of the station.

This indicated value of R is used in solving for the satellite's slant range as in Section IV of the text. It is also used to determine the X_i , Y_i , Z_i station coordinates in Section 5.3.

Normally, azimuth-elevation measurements of a satellite made from a ground station are based upon measurements referenced to a local horizontal plane. This plane is often taken as a tangent to the oblate earth at the station's location. It becomes useful to reference such azimuth-elevation measurements to a plane tangent at the station to a sphere having a radius R . This is done both in trajectory prediction (Section IV) and in preprocessing the azimuth-elevation data prior to orbit determinations.

To relate the azimuth-elevation points to the two tangent planes, we begin by establishing two right-hand coordinate systems, centered on the station. In the first of these, the x - y axes lie in the plane tangent to the oblate earth with the y axis pointing northward and the x axis eastward. In the second system, the x - y' axes lie in the plane tangent to a sphere of radius R . The transformation from one system to the other is a simple one, since rotation of the first set of coordinates about the common x axis moves y into y' as shown in Fig. 25(a).

By inspection of Fig. 25(b), we write the measured azimuth (A) and elevation (E) in terms of the xyz coordinates (referenced to the oblate earth) as

$$x = \rho \cos E \sin A \quad (126)$$

$$y = \rho \cos E \cos A \quad (127)$$

$$z = \rho \sin E \quad (128)$$

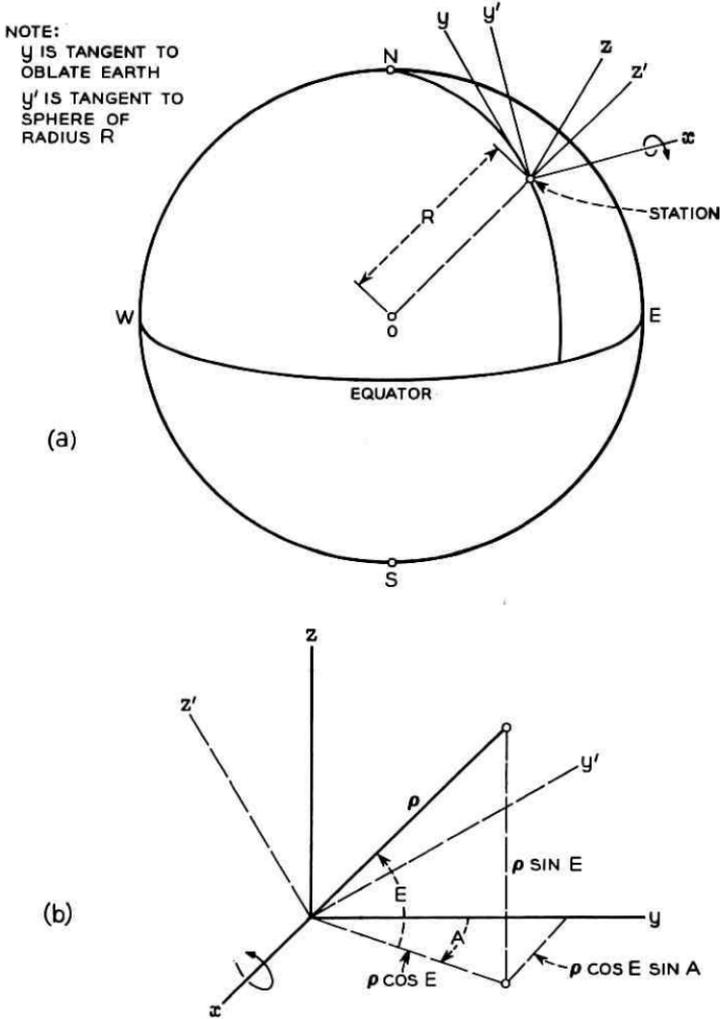


Fig. 25 — Geometry for correcting earth oblateness effects.

where

ρ = the slant range to the satellite.

Next we rotate the xyz coordinates about x through an angle $(\alpha)^*$ equal to the difference between the geodetic and geocentric latitude of the station. This brings y into y' and z into z' where $xy'z'$ is the coordinate

* α = geocentric latitude - geodetic latitude $\approx 0.1962 \sin$ (geodetic latitude).
 A closer approximation is found in Ref. 5.

system for a sphere of radius R . The relationships between y, y' and z, z' are

$$y' = y \cos \alpha + z \sin \alpha \tag{129}$$

$$z' = -y \sin \alpha + z \cos \alpha. \tag{130}$$

Combining (126) through (130) one obtains

$$x = \rho \cos E \sin A \tag{131}$$

$$y' = \rho \cos E \cos A \cos \alpha + \rho \sin E \sin \alpha \tag{132}$$

$$z' = -\rho \cos E \cos A \sin \alpha + \rho \sin E \cos \alpha \tag{133}$$

from which the azimuth (A') and elevation (E') referred to the sphere become

$$E' = \sin^{-1} z' / \rho \tag{134}$$

$$A' = \tan^{-1} x / y'. \tag{135}$$

The procedure is perfectly symmetrical, and if one wants to convert from spherically oriented azimuth-elevation to the oblate earth azimuth-elevation, (131) through (135) are used with a negative α and interchanged primes.

APPENDIX C

The Triangle Ratios m_1 and m_3 as Time Functions

Consider the orbit ellipse lying in the xy plane as shown in Fig. 26. The observed positions of the satellite are expressed in x, y coordinates and time t . If we now define a new time variable $\tau = kt$, the familiar differential equations of motion then become

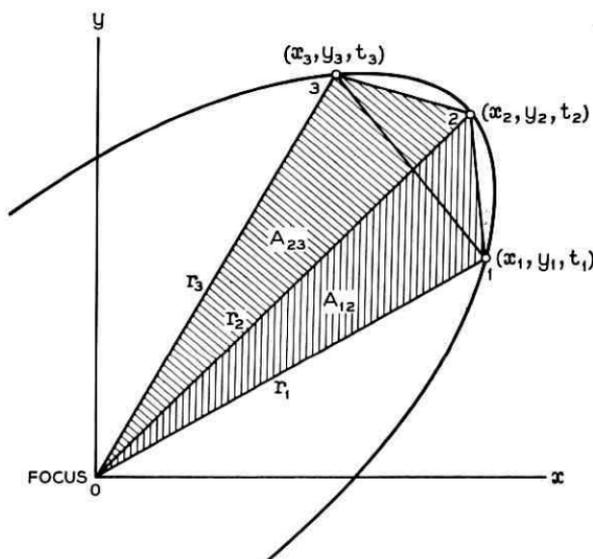
$$\frac{d^2x}{d\tau^2} = -k^2 \frac{x}{r^3} \tag{138}$$

and

$$\frac{d^2y}{d\tau^2} = -k^2 \frac{y}{r^3}. \tag{139}$$

Considering points 1 and 2, we may expand x_1 and y_1 in a Taylor series which generally converges for all reasonable values of τ :

$$x_1 = x_2 + \frac{dx}{d\tau} \tau_3 + \frac{1}{2!} \frac{d^2x}{d\tau^2} \tau_3^2 + \dots \tag{140}$$

Fig. 26 — Orbit triangles in x, y, t space.

$$y_1 = y_2 + \frac{dy}{d\tau} \tau_3 + \frac{1}{2!} \frac{d^2y}{d\tau^2} \tau_3^2 + \dots \quad (141)$$

where the τ derivatives are to be evaluated at the time t_2 of the second data point. We define τ_3 in the text. Similar expressions may of course be written for points 2 and 3. It is possible to substitute for all the derivatives in (140) and (141) higher than the first by successive differentiation of (138) and (139). This permits (140) and (141) to be expressed as

$$x_1 = A_1 x_2 + B_1 \frac{dx}{d\tau} \quad (142)$$

$$y_1 = A_1 y_2 + B_1 \frac{dy}{d\tau} \quad (143)$$

where

$$A_1 = 1 - \frac{1}{2} \left(\frac{\tau_3^2}{r_2^3} + \frac{\tau_3^3}{r_2^4} \frac{dr}{d\tau} \right) + \dots \quad (144)$$

$$B_1 = \tau_3 - \frac{\tau_3^2}{6r_2^3} + \frac{\tau_3^4}{4r_2^4} \frac{dr}{d\tau} + \dots \quad (145)$$

Point x_3, y_3 may be expressed in a manner identical to (142)–(145) with subscript 1 being replaced by 3 throughout and τ_3 becoming τ_2 .

The areas of the triangles formed by r_1 , r_2 , and r_3 and their respective chords are²³

$$A_{12} = \frac{1}{2}(x_1y_2 - y_1x_2) \tag{146}$$

and

$$A_{23} = \frac{1}{2}(x_2y_3 - y_2x_3). \tag{147}$$

If now (142) and (143) are substituted into (146) then,

$$A_{12} = \frac{1}{2}B_1 \left(x_2 \frac{dy}{d\tau} - y_2 \frac{dx}{d\tau} \right). \tag{148}$$

But the right-hand member of (148) is B_1 times the areal velocity* of the satellite, which by Kepler's Third Law is simply

$$A_{12} = \frac{1}{2}B_1k \sqrt{p} \tag{149}$$

where

$$p = a(1 - e^2)$$

k = the Gaussian constant for
earth satellites

$$= 0.07436574 \text{ min}^{-1}.$$

Similar expressions exist for A_{23} which parallel (148) and (149) with subscript 3 substituting for 1. The area enclosed by r_1 , r_3 and the connecting chord differs a bit from either A_{12} or A_{23} and is

$$\begin{aligned} A_{13} &= \frac{1}{2}(x_1y_3 - y_1x_3) \\ &= \frac{1}{2}(A_1B_3 - B_1A_3)k \sqrt{p}. \end{aligned} \tag{150}$$

Now by explicitly expressing the A_i , B_i factors of (144) and (145) in the sector area equations (149) (150), we have

$$A_{12} = \frac{1}{2}k \sqrt{p}\tau_3 \left[1 - \frac{\tau_3^2}{6r_2^3} - \frac{\tau_3^3}{4r_2^4} \frac{dr}{d\tau} + \dots \right] \tag{151}$$

$$A_{23} = \frac{1}{2}k \sqrt{p}\tau_1 \left[1 - \frac{\tau_1^2}{6r_2^3} + \frac{\tau_1^3}{4r_2^4} \frac{dr}{d\tau} + \dots \right] \tag{152}$$

$$A_{13} = \frac{1}{2}k \sqrt{p}\tau_2 \left[1 - \frac{\tau_2^2}{6r_2^3} + \frac{\tau_2(\tau_1^2 - \tau_3)}{4r_2^4} \frac{dr}{d\tau} + \dots \right]. \tag{153}$$

* See Ref. 14; Ref. 24, p. 5; and Ref. 9, p. 26.

The m_1 and m_3 equations follow directly from the above as

$$m_1 = \frac{A_{23}}{A_{13}} \quad (154)$$

$$= \frac{\tau_1}{\tau_2} \left[1 + \frac{\tau_3(\tau_2 + \tau_1)}{6r_2^3} + \frac{\tau_3(\tau_3^2 + \tau_1\tau_3 - \tau_1^2)}{4r_2^4} \frac{dr}{d\tau} + \dots \right]$$

$$m_3 = \frac{A_{12}}{A_{13}} \quad (155)$$

$$= \frac{\tau_3}{\tau_2} \left[1 + \frac{\tau_1(\tau_2 + \tau_3)}{6r_2^3} - \frac{\tau_1(\tau_1^2 + \tau_1\tau_3 - \tau_3^2)}{4r_2^4} \frac{dr}{d\tau} + \dots \right].$$

REFERENCES

1. Spitzer, L., Perturbation of a Satellite's Orbit, J. Brit. Interplanetary Soc., 9, 1950, p. 131.
2. Corneir, L. N., Goodwin, W., and Squires, R. K., Simplified Satellite Predictions from Modified Orbital Elements, IGY Satellite Report Series, No. 7, Natl. Acad. Sci., Natl. Res. Council, Washington, D.C., 1959.
3. Ossanna, J. F., Jr., The Generation of Look Angles from Regularly Available Mean Orbital Elements for 1960 Iota One, Photographic Science and Engr., Nov.-Dec., 1962, pp. 342-350.
4. Moulton, F. R., *An Introduction to Celestial Mechanics*, 2nd rev. ed., The MacMillan Co., New York, 1923.
5. *The American Ephemeris and Nautical Almanac for 1962*, U.S. Govt. Printing Office, 1960.
6. de LaPlace, Marquis, *Mécanique Céleste*, tr. Bowditch, N., Hilliard, Gray, Little and Wilkins, Boston, 1829.
7. Gauss, K. F., *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solum Ambientum*, Hamburg, 1809.
8. Danby, J. M. A., *Fundamentals of Celestial Mechanics*, The MacMillan Co., New York, 1962.
9. Herget, Paul, *The Computation of Orbits*, publ. by author, 1948.
10. Smart, W. M., *Celestial Mechanics*, London, Longmans, Green, 1953, Ch. 5.
11. Claus, A. J., Blackman, R. B., Halline, E. G., and Ridgeway, W. C., III, Orbit Determination and Prediction, and Computer Programs, B.S.T.J., 42, July, 1963, part 2.
12. Notes of the Summer Institute on Dynamical Astronomy at Yale University, July 1959, Yale University Observatory, New Haven, Conn., 1960, pp. 97-100.
13. Blanco, V. M., and McCuskey, S. W., *Basic Physics of the Solar System*, Addison-Wesley Publishing Co., Reading, Mass, 1961.
14. Smart, W. M., *Textbook on Spherical Astronomy*, Cambridge University Press, 1956.
15. *The Explanatory Supplement to the American Ephemeris and Nautical Almanac*, Her Majesty's Stationery Office, London, 1961.
16. King-Hele, D. G., The Earth's Gravitational Potential, in *The Earth Today*, ed. Crook, A. H., and Gaskell, T. F., Royal Astronomical Soc., Interscience Publ., New York.
17. O'Keefe, J. A., Eckels, A., and Squires, R. K., The Gravitational Field of the Earth, *Astrophys. J.*, 64, Sept., 1959.
18. Allen, C. W., *Astrophysical Quantities*, Athlone Press, 1955.
19. Jeffreys, H., *The Earth*, Cambridge University Press, 1952.
20. Lecar, M., Sorenson, J., and Eckels, A., J. Geophys. Res., 64, 1959, pp. 209-216.

21. Merson, R. H., The Motion of a Satellite in an Axisymmetrical Gravitational Field, in *The Earth Today*, ed. Crook, A. H., and Gaskell, T. F., Royal Astronomical Soc., Interscience Publ., New York.
22. Proc. Ninth National Communications Symposium, Oct., 1963, Western Periodicals Co., pp. 146-155.
23. *Handbook of Chemistry and Physics*, 28th ed., Chemical Rubber Publishing Co., Cleveland, O., 1944.
24. McCuskey, S. W., *Introduction to Celestial Mechanics*, Addison-Wesley Publishing Co., Reading, Mass.

Analysis of Varactor Frequency Multipliers for Arbitrary Capacitance Variation and Drive Level

By C. B. BURCKHARDT

(Manuscript received December 23, 1964)

A general analysis of varactor frequency multipliers is given. It applies to multipliers with any rational ratio of output to input frequency, any idler configuration, and any voltage-charge relationship for the diode. A computer program was written for evaluating the results up to a $\times 10$ multiplier. Results are given for the doubler, 1-2-3 tripler, 1-2-4 quadrupler, 1-2-3-4 quadrupler, 1-2-4-5 quintupler, 1-2-4-6 sextupler, and 1-2-4-8 octupler, for various junctions, under conditions of nominal drive as well as of overdriving. Some higher-order multipliers without idlers, which use a high-nonlinearity diode, were computed. Some results are given for a model with varying series resistance.

I. INTRODUCTION

Varactor frequency multipliers have found considerable application for generating microwave signals for receiver local oscillators, parametric amplifier pumps, and other applications.

Penfield and Rafuse¹ have analyzed many multipliers under the assumption that a nominally driven abrupt-junction diode is used. For practical multipliers this assumption often does not hold because the junction is not abrupt and is driven into forward conduction. Greenspan² has analyzed the nominally driven graded-junction doubler, and Davis³ has given an analysis of the overdriven doubler for various capacitance functions.

This paper gives a general analysis of lossy varactor frequency multipliers. The theory applies to multipliers with any rational ratio of output to input frequency, any idler configuration, and any capacitance variation of the diode. A computer program was written for evaluating the results up to a $\times 10$ multiplier, and results are given for the doubler, tripler, quadrupler, quintupler, sextupler and octupler for different drive

levels and different junction capacitance functions. Thus the work reported in this paper can be considered an extension and generalization of the work of the authors mentioned in the previous paragraph.

Circuit losses are neglected in this analysis,* and it is assumed that the only loss occurs in the diode. This restriction was introduced to reduce the number of parameters. Our analysis therefore gives the maximum efficiency that can be obtained with a given diode at a given frequency. Circuit losses usually can be accounted for either by assuming a lower cutoff frequency than that of the diode alone or by computing them separately.

The present analysis differs from that given by Morrison⁴ mainly in two ways:

(a) it assumes a lossy diode, and

(b) it assumes that all the input power is converted to one single output frequency, whereas Ref. 4 is concerned mainly with the maximization of input power under the condition of output power at arbitrary frequencies.

Thus Morrison's results give an upper limit to the power which can be handled by a nonlinear capacitance for prescribed magnitudes of breakdown voltage and forward drive. In addition, he considers the specific examples of graded- and abrupt-junction doublers and triplers, which cases are included in the present results for the limiting lossless condition.

In the analysis in Sections II and III of this paper the diode loss is represented by a constant resistance in series with the variable capacitance. The modification of the analysis due to a variable series resistance is discussed in Section IV. The reader who is interested in the results only is referred to Section V, where all the results are given.

II. ANALYSIS

2.1 Model and Assumptions

The varactor model chosen is a variable capacitance in series with a constant resistance R_s (for R_s variable see Section IV) as shown in Fig. 1. If the applied voltage varies between the contact potential Φ and the breakdown voltage $-V_B$, the diode is said to be fully driven. The overdriven case will also be analyzed. It is assumed that during the period of forward conduction the junction voltage stays clamped to Φ , whereas

* The effect of lossy idlers has been included in the analysis and in the computer program; however, no results have been computed for this case.



Fig. 1 — Varactor model.

the charge varies. Thus the capacitance in our model becomes infinite, and no correction for the rectified forward current is introduced. This model is suitable if the period of forward conduction is several times shorter than the time in which an appreciable number of minority carriers recombine.

It is assumed that currents in the diode flow only at the input, output, and idler frequencies, and that a suitable external circuit prevents other currents from flowing.

2.2 Analysis

The voltage across the variable capacitance can be written as a function of the charge on the capacitance

$$\left(\frac{v_j - \Phi}{V_B - \Phi} \right) = f \left(\frac{q - q_\Phi}{Q_B - q_\Phi} \right) \quad (1)$$

where

- v_j : voltage across the capacitance,
- Φ : contact potential,
- V_B : breakdown voltage,
- q : charge on the capacitance,
- Q_B : charge at breakdown voltage,
- q_Φ : charge at contact potential.

If we introduce the normalized quantities

$$\hat{q} = \frac{q - q_\Phi}{Q_B - q_\Phi}, \quad (2)$$

and

$$\varphi = \frac{v_j - \Phi}{V_B - \Phi}, \quad (3)$$

we can write (1) as

$$\varphi = f(\hat{q}). \quad (4)$$

For the voltage across the diode v_{tot} we have (see Fig. 1)

$$v_{tot} = v_j + R_s i, \quad (5)$$

where i is the current through the diode. Because of (2), (3) and $i = \partial q / \partial t$ we can write (5) as

$$v_{tot} = (V_B - \Phi)f(\hat{q}) + \Phi + R_s(Q_B - q_\Phi)(\partial\hat{q}/\partial t) \quad (6)$$

or

$$v_{tot} - \Phi = (V_B - \Phi)f(\hat{q}) + R_s(Q_B - q_\Phi)(\partial\hat{q}/\partial t). \quad (7)$$

$(v_{tot} - \Phi)$, φ , \hat{q} , and $\partial\hat{q}/\partial t$ can be written as Fourier series

$$\begin{aligned} (v_{tot} - \Phi) &= \sum_{-\infty}^{+\infty} V_k e^{jk\omega_0 t}, \\ \varphi &= \sum_{-\infty}^{+\infty} \varphi_k e^{jk\omega_0 t}, \\ \hat{q} &= \sum Q_k e^{jk\omega_0 t}, \\ \partial\hat{q}/\partial t &= \sum I_k e^{jk\omega_0 t} = \sum jk\omega_0 Q_k e^{jk\omega_0 t}. \end{aligned}$$

The series for \hat{q} and $\partial\hat{q}/\partial t$ have to be summed over all frequencies at which current is flowing in the multiplier.

For the voltage at the load port V_l at frequency $l\omega_0$ we have the relation

$$V_l = (V_B - \Phi)\varphi_l + R_s(Q_B - q_\Phi)I_l = -Z_l(Q_B - q_\Phi)I_l, \quad (8)$$

where Z_l is the load impedance.

$$\begin{aligned} \therefore -Z_l/R_s &= \frac{(V_B - \Phi)}{R_s(Q_B - q_\Phi)} \frac{\varphi_l}{I_l} + 1, \\ -Z_l/R_s &= \kappa \frac{\varphi_l}{jl\omega_0 Q_l} + 1, \end{aligned} \quad (9)$$

where we have introduced the quantity

$$\kappa = \frac{V_B - \Phi}{(Q_B - q_\Phi)R_s}. \quad (10)$$

κ has a simple relation to the familiar cutoff frequency ω_c of Penfield and Rafuse (Ref. 1, p. 86), which may be obtained as follows. The elastance S is defined as

$$S = \partial v_j / \partial q.$$

Because of (2) and (3)

$$S = \frac{\partial \varphi (V_B - \Phi)}{\partial \hat{q} (Q_B - q_\Phi)}.$$

Therefore

$$\begin{aligned} \frac{S}{S_{\max} - S_{\min}} &= \frac{\partial \varphi}{\partial \hat{q}} \frac{(V_B - \Phi)}{(Q_R - q_\Phi) R_S} \frac{R_S}{S_{\max} - S_{\min}} \\ &= \frac{\partial \varphi}{\partial \hat{q}} \frac{\kappa}{\omega_c}, \end{aligned} \quad (11)$$

where S_{\max} is the maximum value of the elastance, S_{\min} is the minimum value of the elastance and

$$\omega_c = \frac{S_{\max} - S_{\min}}{R_S}$$

is the cutoff frequency defined by Penfield and Rafuse.

If we set $S = S_{\max}$ in (11) we obtain

$$\frac{S_{\max}}{S_{\max} - S_{\min}} = \frac{\partial \varphi}{\partial \hat{q}} \Big|_{\hat{q}=1} \cdot \frac{\kappa}{\omega_c}$$

and

$$\kappa = \frac{\omega_c}{\frac{\partial \varphi}{\partial \hat{q}} \Big|_{\hat{q}=1}} \cdot \frac{S_{\max}}{S_{\max} - S_{\min}}. \quad (12)$$

This relation between κ and ω_c is dependent on the nonlinearity of the diode. E.g., if we assume an abrupt junction with $\varphi = \hat{q}^2$ and $S_{\min} = 0$ we obtain $\kappa = 0.5 \omega_c$.

We now want to solve (9) for the real and imaginary parts of Z_l . Fig. 2 shows the quantities of interest plotted in the complex plane. The input current I_{in} is assumed to be real and positive. With the phase angles as defined we obtain from (9)

$$-R_l/R_S = \kappa \frac{|\varphi_l|}{|Q_l|} \frac{1}{l\omega_0} \sin(-\alpha + \beta_l) + 1.$$

If we solve this for $|Q_l|$ we obtain

$$|Q_l| = \frac{\kappa |\varphi_l| \sin(-\alpha + \beta_l)}{(-R_l/R_S - 1) l\omega_0}. \quad (13)$$

For the imaginary part X_l we obtain

$$X_l/R_S = \kappa \frac{|\varphi_l|}{|Q_l|} \frac{1}{l\omega_0} \cos(-\alpha + \beta_l). \quad (14)$$

If we define an effective output elastance S_{effout} according to $S_{\text{effout}}/\omega_0 l = X_l$ we obtain

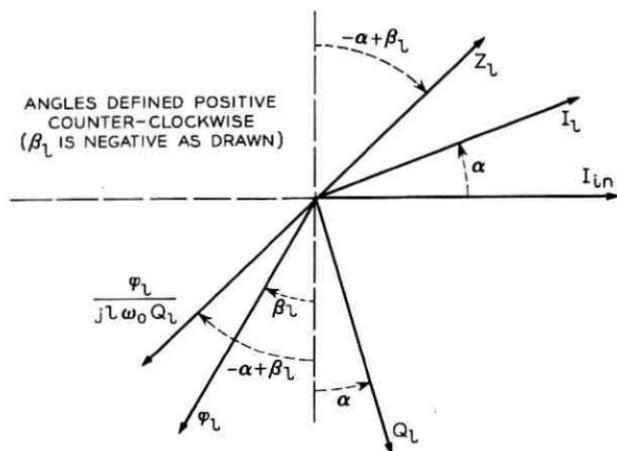


Fig. 2 — Output quantities in complex plane.

$$\frac{S_{\text{effout}}}{S_{\text{max}} - S_{\text{min}}} = \frac{\omega_0 l X_l}{S_{\text{max}} - S_{\text{min}}} = \frac{R_s \kappa}{S_{\text{max}} - S_{\text{min}}} \frac{|\varphi_l|}{|Q_l|} \cos(-\alpha + \beta_l)$$

$$\frac{S_{\text{effout}}}{S_{\text{max}} - S_{\text{min}}} = \frac{\kappa}{\omega_c} \frac{|\varphi_l|}{|Q_l|} \cos(-\alpha + \beta_l). \quad (15)$$

Idler ports: The idler impedance Z_i at the frequency $i\omega_0$ is

$$-Z_i/R_s = \kappa \frac{\varphi_i}{j \cdot i\omega_0 Q_i} + 1. \quad (16)$$

We now assume that all the idler ports are tuned,* i.e., we assume the idler currents to be in phase with I_{in} . We then obtain

$$-R_i/R_s = \kappa \frac{|\varphi_i|}{|Q_i|} \frac{1}{i\omega_0} \sin \beta_i + 1 \quad (17)$$

and

$$|Q_i| = \frac{\kappa |\varphi_i| \sin \beta_i}{(-R_i/R_s - 1) i\omega_0}, \quad (18)$$

where R_i is the idler resistance and β_i is defined analogous to β_l . For the effective idler elastance S_{effidl} we obtain

* Penfield and Rafuse¹ find that tuning the output circuit gives near optimum efficiency for the nominally driven abrupt-junction doubler. Davis² obtains the same result for a variety of nominally driven as well as overdriven abrupt- and graded-junction doublers. Judging from these results, it was felt that tuning the idlers leads to near optimum efficiency, and the above-mentioned restriction was introduced.

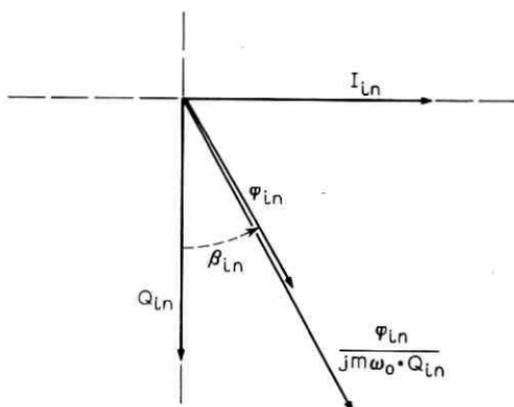


Fig. 3 — Input quantities in complex plane.

$$\frac{S_{\text{eff}i1}}{S_{\text{max}} - S_{\text{min}}} = \frac{\kappa}{\omega_c} \frac{|\varphi_i|}{|Q_i|} \cos \beta_i. \quad (19)$$

Input port: The input impedance Z_{in} at $m\omega_0$ is

$$Z_{\text{in}} = \frac{\text{input voltage}}{\text{input current}}$$

$$Z_{\text{in}} = \frac{(V_B - \Phi)\varphi_{\text{in}}}{I_{\text{in}}(Q_B - q\Phi)} + R_S,$$

$$Z_{\text{in}}/R_S = \frac{\kappa\varphi_{\text{in}}}{jm\omega_0 Q_{\text{in}}} + 1. \quad (20)$$

The input quantities are plotted in Fig. 3. For the input resistance R_{in} we obtain

$$R_{\text{in}}/R_S = \frac{\kappa |\varphi_{\text{in}}|}{m\omega_0 |Q_{\text{in}}|} \sin \beta_{\text{in}} + 1 \quad (21)$$

and for the imaginary part X_{in}

$$X_{\text{in}}/R_S = -\frac{\kappa |\varphi_{\text{in}}|}{m\omega_0 |Q_{\text{in}}|} \cos \beta_{\text{in}}.$$

$$\frac{S_{\text{eff}in}}{S_{\text{max}} - S_{\text{min}}} = \frac{\kappa}{\omega_c} \frac{|\varphi_{\text{in}}|}{|Q_{\text{in}}|} \cos \beta_{\text{in}}. \quad (22)$$

Power relations and efficiency: The input power P_{in} is

$$P_{\text{in}} = 2 |I_{\text{in}}|^2 R_{\text{in}} (Q_B - q\Phi)^2$$

(I_{in} is a half-amplitude). Introducing the normalization power P_{norm}

$$P_{norm} = \frac{(V_B - \Phi)^2}{R_S}$$

we can write

$$P_{in}/P_{norm} = 2m^2\omega_0^2 |Q_{in}|^2 R_{in} \frac{(Q_B - q\Phi)^2}{(V_B - \Phi)^2} R_S,$$

$$P_{in}/P_{norm} = \frac{2m^2\omega_0^2 |Q_{in}|^2}{\kappa^2} (R_{in}/R_S). \quad (23)$$

Similarly, we obtain for the normalized output power P_{out}/P_{norm}

$$P_{out}/P_{norm} = \frac{2l^2\omega_0^2 |Q_l|^2}{\kappa^2} (R_l/R_S). \quad (24)$$

For the normalized power dissipated in the idler resistances P_{diss1}/P_{norm} we obtain

$$P_{diss1}/P_{norm} = 2 \frac{\omega_0^2}{\kappa^2} \sum_i i^2 |Q_i|^2 (R_i/R_S), \quad (25)$$

where the sum has to be extended over all frequencies at which there are idlers. The normalized power dissipated in the series resistance R_S of the diode, P_{diss2}/P_{norm} is

$$P_{diss2}/P_{norm} = 2 \frac{\omega_0^2}{\kappa^2} \sum_k k^2 |Q_k|^2, \quad (26)$$

with the sum extended over all frequencies at which there are currents flowing. The total power dissipated P_{diss} is

$$P_{diss} = P_{diss1} + P_{diss2}. \quad (27)$$

For the efficiency ϵ we obtain

$$\epsilon = 1 - (P_{diss}/P_{in}) \quad (28)$$

$$= P_{out}/P_{in}. \quad (29)$$

Bias voltage: The normalized bias voltage

$$\varphi_0 = \left(\frac{V_0 - \Phi}{V_B - \Phi} \right)$$

is equal to the constant coefficient in the Fourier series for φ .

Evidently the foregoing analysis holds for any integrally related in-

put, output, and idler frequencies and therefore can be applied to any corresponding multiplier configuration.

III. TECHNIQUE OF SOLUTION AND OUTLINE OF THE COMPUTER PROGRAM

For a multiplier the following quantities are usually given physically:

- (a) input, output, and idler frequencies,
- (b) the voltage-charge relation of the diode and its series resistance R_s , and
- (c) the idler resistances R_i .

The following quantities will be treated as parameters, i.e., they will be set constant for one evaluation and will be varied later in order to optimize a certain quantity, usually the efficiency. These parameters are:

- (d) the minimum charge Q_{\min} and the maximum charge Q_{\max} between which the diode is driven,
- (e) the load resistance R_l of the multiplier, and
- (f) the angle of output detuning α .*

The formulas of Section 2.2 now have to be evaluated subject to these constraints. The solution proceeds in a way suggested physically. A current is applied at the input port. This current gives rise to certain voltages at the idler and output ports which in turn cause certain currents at these ports. The calculation of the voltages at the idler and output ports is repeated assuming the diode to be driven by the sum of the currents computed previously. The computation is repeated until the values of all the currents have reached their asymptotic values.

Fig. 4 shows a simplified block diagram of the computer program. First, a value for the input charge coefficient $|Q_{\text{in}}|$ and a value for the load resistance R_l are assumed. The normalized charge \hat{q} is computed at n equidistant points lying in one period of the first harmonic. The maximum and minimum values of \hat{q} are determined† and compared with the values of \hat{q}_{\max} and \hat{q}_{\min} prescribed. The values of $|Q_k|$ are corrected‡ in order to give a variation of \hat{q} between \hat{q}_{\max} and \hat{q}_{\min} and \hat{q} is computed again. From this and the voltage charge law of the diode (4) the normalized voltage φ is computed. The Fourier coefficients φ_k of φ are then evaluated by a subprogram as described in Ref. 5. From the values of

* For all the computations in this paper $\alpha = 0^\circ$ for the same reason given in the footnote to Section 2.2.

† The maximum and minimum values of the n values of \hat{q} are obtained by a sorting routine. The maximum and minimum of \hat{q} are found by quadratic interpolation.

‡ Shift of $|Q_0|$; compression or expansion of $|Q_k|$, $k \geq 1$.

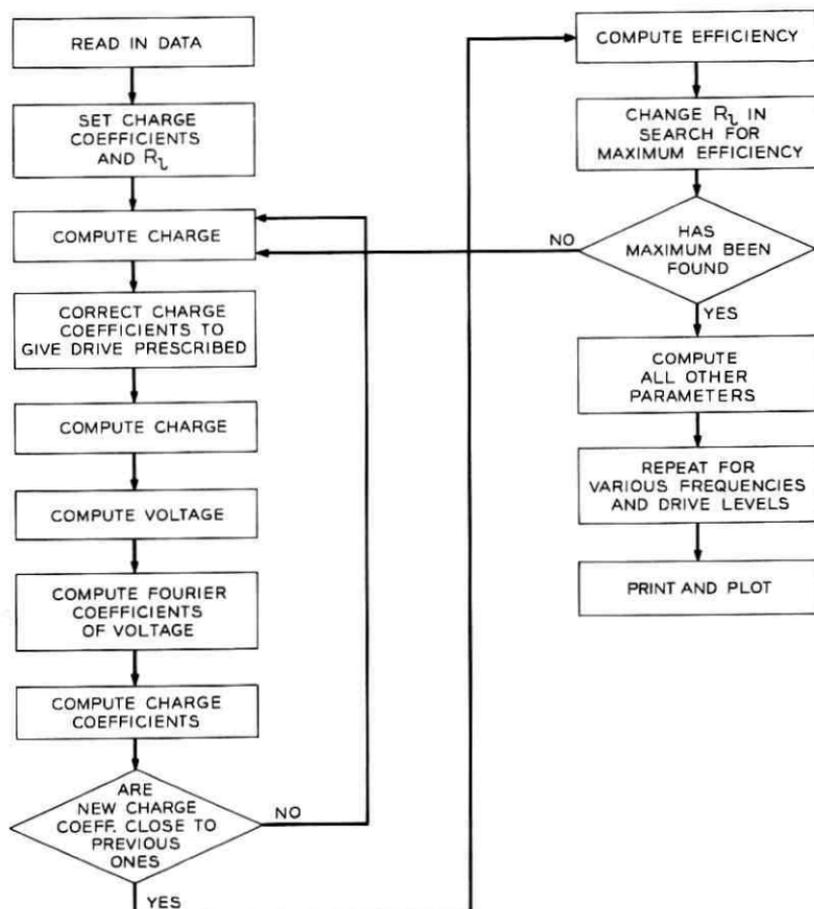


Fig. 4 — Simplified flow diagram of computer program.

φ_k and (13) and (18) new values can be computed for the charge coefficients. The computation is repeated until all the charge coefficients computed in step $(n + 1)$ are very close to those computed in step n . The computation as described does not converge if the multiplier has idlers but rather gives an oscillatory solution for the charge coefficients. It can be stabilized in the following way. Suppose $|Q_i|_{n+1}'$ is the idler charge coefficient computed in step $(n + 1)$ according to (18). One then computes $|Q_i|_{n+1}$ as a weighted average of $|Q_i|_{n+1}'$ and the charge coefficient $|Q_i|_n$ computed in the previous step,

$$|Q_i|_{n+1} = \frac{w |Q_i|_n + |Q_i|_{n+1}'}{w + 1}. \quad (30)$$

The problem now is to find a suitable value for w . If w is too small the solution does not converge; if it is too big the convergence is very slow. Numerous experiments yielded the following (heuristic) formula for w ,

$$w = \frac{0.15(m+1)\omega_c \sqrt{n}}{\omega_0} \quad (31)$$

where m is the number of idlers at higher frequency than the idler frequency at which $|Q_i|$ is being computed; n is the total number of idlers in the multiplier. Equation (31) gave a good initial choice for the values of w . In some cases these values had to be modified slightly, either to speed up convergence or to prevent oscillation. The charge coefficient $|Q_l|$ at the output frequency can be computed according to (13).

When the final values of the $|Q_k|$'s have been reached, the efficiency is computed using (28). For small values of ϵ , (29) is used because it gives more accurate results. R_l is then changed, and the computation is repeated. R_l is changed until a point of maximum efficiency is reached. The optimum value of R_l is then found by quadratic interpolation, and the computation is carried out a last time with this optimum value. After this the values of S_{eff} are computed. The computation is carried out for a number of ratios ω_0/ω_c and a number of drive levels.

The accuracy of the computation was checked by computing the nominally driven abrupt-junction doubler, 1-2-3 tripler, 1-2-4-8 octupler, and the graded-junction doubler. All the results agreed with known results^{1,2} within plotting accuracy.

IV. EFFECT OF A VARYING SERIES RESISTANCE R_s

For point contact varactors the model of Fig. 1 is a good representation. For epitaxial varactors, however, R_s depends significantly on bias voltage,⁶ and R_s should be considered a function of \hat{q} in our computation. Equation (7) then becomes

$$v_{\text{tot}} - \Phi = (V_B - \Phi)f(\hat{q}) + (Q_B - q_\Phi)(\partial\hat{q}/\partial t)R_s(\hat{q}). \quad (32)$$

Because (32) seemed to be too difficult to be solved exactly a *semi-rigid* approach was made. From the results for the constant series resistance model we know that, for low and moderately high frequencies, input and load resistances of a multiplier are several times bigger than R_s . We conclude that the second term in (32) is small compared with the first one. We normalize $R_s(\hat{q})$ with respect to some value R_{s0} chosen arbitrarily. The computation proceeds as in Section II and *only* the computation of the power dissipated P_{diss2} in $R_s(\hat{q})$ is *modified* as follows

$$P_{\text{diss2}} = \frac{1}{2\pi} \int_{\omega_0 t=0}^{2\pi} R_S(\hat{q}) i^2 d(\omega_0 t), \quad (33)$$

where i is the instantaneous current in R_S . With

$$\hat{q} = \sum 2 |Q_k| \sin k\omega_0 t$$

one obtains

$$P_{\text{diss2}}/P_{\text{norm}} = \frac{\omega_0^2}{2\pi k^2} \int_{\omega_0 t=0}^{2\pi} \frac{R_S(\hat{q})}{R_{S0}} [\sum 2 |Q_k| |k \cos k\omega_0 t|^2] d(\omega_0 t). \quad (34)$$

Equation (34) was substituted for (26) in the program. The integral in (34) was evaluated numerically.

V. RESULTS

5.1 Computations with R_S Constant

The results reported in this section were computed for a constant series resistance model and a voltage-charge relation of the form

$$\frac{v - \Phi}{V_B - \Phi} = \left(\frac{q - q_\Phi}{Q_B - q_\Phi} \right)^{1/(1-\gamma)}. \quad (35)$$

Values of $\gamma = 0.5$ (abrupt), $\gamma = 0.4$, $\gamma = 0.333$ (graded), and $\gamma = 0$ were used.* Results were computed for different drive levels.

Plots of all the quantities versus ω_0/ω_c were obtained using the IBM 7094 computer and the SC-4020 microfilm printer. Lack of space prohibits the reproduction of these plots. Instead, the constants occurring in the low-frequency approximations used by Rafuse⁷ will be given. These low-frequency approximations are

$$\epsilon = \exp(-\alpha \omega_{\text{out}}/\omega_c), \dagger$$

$$P_{\text{out}}/P_{\text{norm}} = \beta(\omega_0/\omega_c),$$

$$R_{\text{in}}/R_S = A(\omega_c/\omega_0),$$

$$R_l/R_S = B(\omega_c/\omega_0).$$

The normalized elastances as well as the normalized bias voltage do not depend on ω_0/ω_c at low frequencies. For the tripler and higher-order multipliers the output power versus frequency shows a maximum. This value, designated $P_{\text{max}}/P_{\text{norm}}$, and the corresponding frequency, design-

* $\gamma = 0$ corresponds to a step in elastance at $\hat{q} = 0$. $S = 0$ for $\hat{q} \leq 0$ and $S =$ a constant for $\hat{q} \geq 0$.

† α was computed from the value of ϵ at $\omega_0 = 10^{-2}\omega_c$.

TABLE I — DOUBLER

Drive →	$\gamma = 0.0$		$\gamma = 0.333$			$\gamma = 0.4$			$\gamma = 0.5$	
	1.5	2.0	1.0	1.3	1.6	1.0	1.3	1.6	1.3	1.6
α	6.7	4.7	12.6	8.0	6.9	11.1	8.0	7.2	8.3	8.3
β	0.0222	0.0626	0.0118	0.0329	0.0587	0.0168	0.0406	0.0678	0.0556	0.0835
A	0.117	0.213	0.0636	0.101	0.126	0.0730	0.102	0.118	0.0980	0.0977
B	0.204	0.211	0.0976	0.158	0.172	0.112	0.157	0.161	0.151	0.151
S_{01}/S_{\max}	0.73	0.50	0.68	0.52	0.40	0.61	0.45	0.35	0.37	0.28
S_{02}/S_{\max}	0.60	0.50	0.66	0.48	0.41	0.59	0.44	0.38	0.40	0.34
$V_{0\text{norm}}$	0.35	0.25	0.41	0.33	0.27	0.39	0.31	0.26	0.28	0.24

nated $\omega_{0\max}/\omega_c$, also will be given. These values will help to give a better reconstruction of the actual curves.

The results are contained in Tables I–VIII. The notation 1-2-3-4 quadrupler means that there are idlers at $2\omega_0$ and $3\omega_0$. A 1-8 octupler has no idlers.

The drive is defined as

$$\text{drive} = \frac{Q_{\max} - Q_{\min}}{Q_B - q\Phi} \quad (36)$$

Thus drive = 1 corresponds to nominal drive and drive > 1 corresponds to overdriving the junction. For all drive levels it was assumed that the junction was driven up to the charge corresponding to breakdown in the reverse direction.

Tables I–VII show that the power handling capability as well as the efficiency can be increased by overdriving the diode. This increase in efficiency is most pronounced for diodes having low values of γ . For the same cutoff frequency the highest efficiency is obtained for a junction with $\gamma = 0$ and drive = 2.0, as is seen in Tables I and III for the doubler and the 1-2-4 quadrupler.

Sometimes the question is asked: What efficiency can one obtain for a multiplier without idlers and a high-nonlinearity diode, and do idlers improve the performance of such a multiplier? To answer this question, a quadrupler, sextupler, and octupler without any idlers were computed for a diode with $\gamma = 0$. The highest efficiency was obtained for drive = 2.0 for all these multipliers. The results of Table VIII show that the efficiencies of these multipliers are about as high as those of the nominally driven abrupt-junction 1-2-3-4 quadrupler,⁸ 1-2-4-6 sextupler, and 1-2-4-8 octupler.¹ The power handling capability of the 1-4 quadrupler is lower than that of the abrupt-junction 1-2-3-4 quadrupler, and it compares even less favorably for the sextupler and octupler. Compari-

TABLE IV — 1-2-3-4 QUADRUPLER

Drive →	$\gamma = 0.333$			$\gamma = 0.5$	
	1.0	1.3	1.6	1.3	1.6
α	14.1	8.9	8.1	9.4	9.7
β	0.0094	0.0260	0.0438	0.0439	0.0647
P_{\max}/P_{norm}	$1.1 \cdot 10^{-4}$	$4.8 \cdot 10^{-4}$	$8.2 \cdot 10^{-4}$	$7.4 \cdot 10^{-4}$	$1.1 \cdot 10^{-3}$
ω_{\max}/ω_c	$3.0 \cdot 10^{-2}$	$6.4 \cdot 10^{-2}$	$6.3 \cdot 10^{-2}$	$5.2 \cdot 10^{-2}$	$6.0 \cdot 10^{-2}$
A	0.0719	0.118	0.155	0.120	0.122
B	0.0489	0.0797	0.0927	0.0748	0.0729
S_{01}/S_{\max}	0.69	0.55	0.40	0.36	0.25
S_{02}/S_{\max}	0.66	0.48	0.41	0.39	0.34
S_{03}/S_{\max}	0.67	0.51	0.42	0.38	0.31
S_{04}/S_{\max}	0.67	0.50	0.40	0.38	0.30
V_{norm}	0.40	0.30	0.23	0.25	0.20

TABLE V — 1-2-4-5 QUINTUPLER

Drive →	$\gamma = 0.333$			$\gamma = 0.5$	
	1.0	1.3	1.6	1.3	1.6
α	21.4	14.5	14.8	15.8	16.6
β	0.0072	0.0198	0.0310	0.0326	0.0470
P_{\max}/P_{norm}	$4.2 \cdot 10^{-5}$	$1.6 \cdot 10^{-4}$	$2.4 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	$3.4 \cdot 10^{-4}$
ω_{\max}/ω_c	$1.4 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	$2.0 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$
A	0.104	0.170	0.203	0.167	0.163
B	0.0315	0.0524	0.0592	0.0485	0.0470
S_{01}/S_{\max}	0.69	0.54	0.39	0.36	0.25
S_{02}/S_{\max}	0.69	0.54	0.40	0.36	0.26
S_{04}/S_{\max}	0.68	0.53	0.41	0.37	0.28
S_{05}/S_{\max}	0.67	0.49	0.40	0.38	0.32
V_{norm}	0.40	0.29	0.23	0.24	0.19

TABLE VI — 1-2-4-6 SEXTUPLER

Drive →	$\gamma = 0.333$			$\gamma = 0.5$	
	1.0	1.3	1.6	1.3	1.6
α	19.6	13.0	11.3	13.4	13.7
β	0.0086	0.0239	0.0419	0.0405	0.0598
P_{\max}/P_{norm}	$4.1 \cdot 10^{-5}$	$1.7 \cdot 10^{-4}$	$3.3 \cdot 10^{-4}$	$2.7 \cdot 10^{-4}$	$4.0 \cdot 10^{-4}$
ω_{\max}/ω_c	$1.4 \cdot 10^{-2}$	$2.3 \cdot 10^{-2}$	$2.0 \cdot 10^{-2}$	$2.5 \cdot 10^{-2}$	$2.3 \cdot 10^{-2}$
A	0.0877	0.145	0.177	0.140	0.140
B	0.0179	0.0290	0.0314	0.0271	0.0259
S_{01}/S_{\max}	0.69	0.54	0.40	0.36	0.26
S_{02}/S_{\max}	0.68	0.52	0.40	0.37	0.28
S_{04}/S_{\max}	0.69	0.56	0.41	0.36	0.24
S_{06}/S_{\max}	0.68	0.53	0.41	0.37	0.28
V_{norm}	0.40	0.32	0.26	0.27	0.22

TABLE VII — 1-2-4-8 OCTUPLER

Drive →	$\gamma = 0.333$			$\gamma = 0.5$	
	1.0	1.3	1.6	1.3	1.6
α	28.4	17.8	13.9	17.7	17.5
β	0.0071	0.0205	0.0380	0.0355	0.0537
P_{\max}/P_{norm}	$1.7 \cdot 10^{-5}$	$7.2 \cdot 10^{-5}$	$1.6 \cdot 10^{-4}$	$1.2 \cdot 10^{-4}$	$1.9 \cdot 10^{-4}$
$\omega_{0\max}/\omega_c$	$5.0 \cdot 10^{-3}$	$1.1 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$8.0 \cdot 10^{-3}$	$9.0 \cdot 10^{-3}$
A	0.0795	0.129	0.153	0.125	0.124
B	0.0156	0.0220	0.0255	0.0217	0.0212
S_{01}/S_{\max}	0.68	0.53	0.40	0.37	0.27
S_{02}/S_{\max}	0.68	0.53	0.41	0.37	0.28
S_{04}/S_{\max}	0.68	0.52	0.41	0.37	0.28
S_{08}/S_{\max}	0.68	0.50	0.39	0.37	0.30
$V_{0\text{norm}}$	0.41	0.33	0.28	0.28	0.24

son of Table VIII and Table III shows that an idler at $2\omega_0$ improves both the efficiency and power handling capability of the quadrupler using a diode with $\gamma = 0$.

5.2 Computations with R_s Variable

Computations were carried out for the doubler and 1-2-3 tripler with abrupt and graded junctions. The series resistance was assumed to be that of the semiconductor wafer *only*: i.e., any contribution due to contact resistance etc. was neglected. The normalization resistance R_{s0} was arbitrarily chosen at $\hat{q} = 0.5$.

With this normalization we obtain (Ref. 1, pp. 515-6)

$$R_s/R_{s0} = 2(1 - \hat{q}) \quad (37)$$

for the abrupt junction and

$$R_s/R_{s0} = 2.91 \ln \hat{q}^{-\frac{1}{2}} \quad (38)$$

TABLE VIII — MULTIPLIERS WITHOUT IDLERS
($\gamma = 0$, Drive = 2.0)

Multiplier	1-4	1-6	1-8
α	11.8	17.6	21.7
β	0.0144	0.0063	0.0034
P_{\max}/P_{norm}	$2.2 \cdot 10^{-4}$	$4.1 \cdot 10^{-5}$	$1.3 \cdot 10^{-5}$
$\omega_{0\max}/\omega_c$	$1.0 \cdot 10^{-1*}$	$1.0 \cdot 10^{-1*}$	$3.0 \cdot 10^{-2*}$
A	0.0415	0.0175	0.0098
B	0.0430	0.0189	0.0106
S_{01}/S_{\max}	0.50	0.50	0.50
$S_{0\text{out}}/S_{\max}$	0.50	0.50	0.50
$V_{0\text{norm}}$	0.27	0.28	0.29

* Broad maxima.

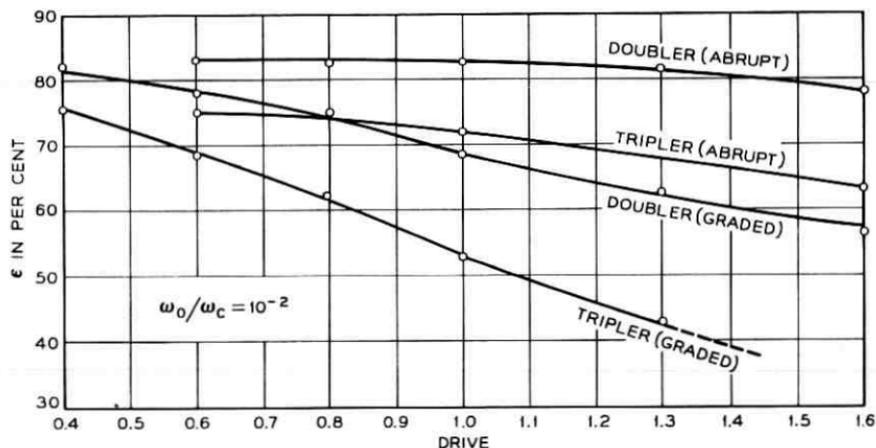


Fig. 5 — Efficiency of variable R_s multipliers vs drive level.

for the graded junction. Equation (38) has a pole at $\hat{q} = 0$ which in reality does not occur because of the nonzero intrinsic conductivity of the material. Arbitrarily, it was assumed that R_s/R_{s0} varied according to (38) for $\hat{q} \geq 0.005$ and had the value corresponding to $\hat{q} = 0.005$ for $\hat{q} \leq 0.005$.

Fig. 5 shows the efficiency of the abrupt- and graded-junction doublers and triplers as a function of the drive level at $\omega_0/\omega_c = 10^{-2}$. The efficiencies of the abrupt-junction multipliers decrease slightly; those of the graded-junction multipliers decrease more rapidly with increasing drive level. These results are in contrast to the results for the model with constant series resistance. For the graded junction the values of efficiency for drive > 1 , of course, depend on the value of \hat{q} below which R_s/R_{s0} is assumed to be constant. The values for drive < 1 , however, do not.

The complete tables for these multipliers are not included here, because all the values can be deduced from the results of the constant series computation as follows:

All the values *except efficiency* are identical to the results for the con-

TABLE IX — VALUES OF k FOR VARIOUS DRIVE VALUES

Drive	1.0	1.3	1.6
Abrupt-junction doubler	1.0	1.2	1.5
Abrupt-junction tripler	1.0	—	1.5
Graded-junction doubler	1.5	3.0	4.0
Graded-junction tripler	1.5	2.8	4.0

stant series resistance computation within 10 per cent. The values for the efficiency can be read from the tables for the constant R_s model. To do so one multiplies the input frequency ω_0 of the variable R_s multiplier by k and computes the efficiency at the frequency

$$\omega_0' = k\omega_0$$

using the value of α given in the constant R_s tables. These values agree with the computed values within 3 per cent for efficiencies larger than 50 per cent.

The value of k is given in Table IX.

Needless to say, the correspondence given above applies to our particular definition of R_{s0} only.

VI. ACKNOWLEDGMENT

The author is indebted to J. W. Gewartowski for suggesting the subject of this investigation and for many helpful discussions.

REFERENCES

1. Penfield, P., and Rafuse, R. P., *Varactor Applications*, M.I.T. Press, 1962, pp. 329-412.
2. Greenspan, M., The Graded-Junction Varactor Frequency Doubler, S. M. Thesis, Dept. Elect. Engrng., M.I.T., June, 1962.
3. Davis, J. A., The Forward-Driven Varactor Frequency Doubler, S.M. Thesis, Dept. Elect. Engrng., M.I.T., May, 1963.
4. Morrison, J. A., Maximization of the Fundamental Power in Nonlinear Capacitance Diodes, B.S.T.J., 41, March, 1962, pp. 677-721.
5. Ralston, A., and Wilf, H. S., *Mathematical Methods for Digital Computers*, J. Wiley, 1960, pp. 258-262.
6. Lee, T. P., Evaluation of Voltage-Dependent Series Resistance of Epitaxial Varactor Diodes at Microwave Frequencies, 1964 Electron Devices Meeting, October 29-31, Washington, D. C.
7. Rafuse, R. P., Recent Developments in Parametric Multipliers, Proc. Nat'l. Electronics Conference, October, 1963, 19, pp. 461-470.
8. Diamond, B. L., Some Results for Higher-Order Varactor Frequency Multipliers, Report 47G-5, Lincoln Lab, M.I.T., March 21, 1963, p. 48.

Acoustic Beam Probing Using Optical Techniques

By M. G. COHEN and E. I. GORDON

(Manuscript received January 8, 1965)

In Section I it is demonstrated that the amplitude of the light deflected or scattered by an advancing sinusoidal acoustic wave, as a function of the angle between the direction of light propagation and the acoustic wavefront, is proportional to the Fourier transform of the amplitude distribution of the acoustic wave in the plane of the wavefront. Studying the angular dependence of the optical-acoustic interaction accurately and directly determines the angular distribution or far-field diffraction pattern of the acoustic beam and incidentally determines the angular response of the acoustic transducer producing the beam. The angular resolution equals the angular spread in the probing light beam. Experiments illustrating and verifying the technique are described.

In Section II the effect of volume acoustic loss is determined. It is shown that loss does not change the considerations of Section I apart from a slight reduction in angular resolution unless the decay distance is comparable to the acoustic wavelength. The loss parameter does introduce a maximum usable acoustic beam width for the interaction (coherence width). In addition, techniques for determining the acoustic loss are described. Particular attention is given to the near- and far-field energy distribution of the scattered light beam. It is shown that the far-field distribution is Lorentzian only under special circumstances. Consideration is given to probing beams with rectangular and Gaussian intensity distributions. Edge effects are taken into account, and it is shown that these can make important contributions to the line shape as well as lead to errors in the interpretation of phonon lifetimes from scattering experiments. Experiments confirming the results are described.

INTRODUCTION

It is well known that acoustic waves in transparent materials can be used to deflect or scatter light beams.^{1,2} As a result, a great deal can be learned about the energy distribution in the acoustic beam by studying the angular and positional dependence of the optical-acoustic interaction.

The paper is divided into two parts. Section I is devoted to the theoretical and experimental demonstration of the fact that the amplitude of the light deflected by an advancing sinusoidal acoustic wave, as a function of the angle between the direction of light propagation and the acoustic wavefront, is proportional to the Fourier transform of the amplitude distribution of the acoustic wave in the plane of the wavefront. Thus the angular dependence of the optical-acoustic interaction accurately and directly measures the angular distribution of the acoustic energy. Stated another way, a study of the total power in the scattered light beam as a function of the angle of the light beam relative to the acoustic beam yields directly the far-field or Fraunhofer diffraction pattern of the acoustic beam.

The power in the deflected light beam measures the acoustic intensity at the position of the light beam. Absolute determination of the acoustic intensity can be made if the photoelastic constants for the medium are known. This technique is thereby capable of providing more information about the acoustic beam than can be obtained with conventional pulse-echo techniques or acoustic probes, such as described by Fitch and Dean,³ which must be used at a boundary of the acoustic transmission medium. In particular, volume acoustic loss can be determined directly. The precise direction and phase velocity of the acoustic wave can also be determined unequivocally.

In Section 1.1, the theory of the optical-acoustic interaction in the absence of volume loss is sketched. Experiments verifying the basic concepts are described in Section 1.2.

In Section II the case of finite volume loss is considered. Particular attention is given to techniques for determining the volume loss both by probing the acoustic beam along its propagation path and by observing the far-field diffraction of the scattered light beam.

I. OPTICAL-ACOUSTIC INTERACTION IN THE ABSENCE OF VOLUME LOSS

1.1 *Theory*

The geometry of the interaction is defined in Fig. 1. The acoustic wave is propagating approximately along the x axis, and the light

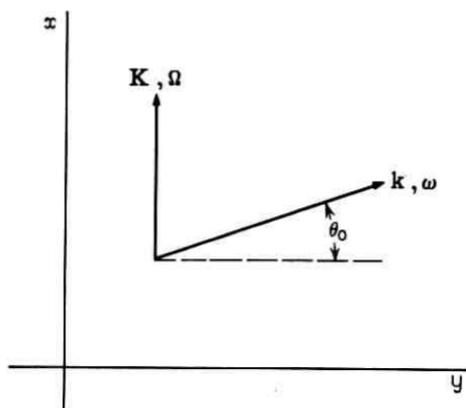


Fig. 1 — Geometry of the interaction region. The z axis is out of the plane of the paper.

beam in the x - y plane. It is assumed there is no z variation. For light with a given polarization, the photoelastic effect produces a variation in the dielectric constant which is proportional to the amplitude of the acoustic wave. The amplitude of the acoustic wave is such that the dielectric constant of the medium ϵ has a variation $\Delta\epsilon$ given by

$$\Delta\epsilon(x,y,t) = [\Delta\epsilon(y)]_c \cos(\Omega t - Kx) + [\Delta\epsilon(y)]_s \sin(\Omega t - Kx) \quad (1)$$

in which Ω is the acoustic angular frequency and K the acoustic propagation constant ($K = \Omega/v$ with v the acoustic velocity). It can be shown that the optical field E of angular frequency ω is described by the wave equation¹

$$\partial^2 E / \partial x^2 + \partial^2 E / \partial y^2 - c'^{-2} \partial^2 (1 + \Delta\epsilon/\epsilon) E / \partial t^2 = 0 \quad (2)$$

in which c' is the light velocity in the medium. In the absence of a perturbation in the dielectric constant, E is assumed to be a plane wave with propagation constant $k = \omega/c'$. As in Ref. 1, the perturbed E is expanded in a set of plane waves appropriate to the grating orders associated with a periodic index variation,

$$E(x,y,t) = \sum_{l=-\infty}^{+\infty} V_l(y) \exp i[(\omega + l\Omega)t - (k \sin \theta_0 + lK)x - ky \cos \theta_0] + \text{complex conjugate.} \quad (3)$$

The quantity $V_l(y)$ can be identified as the amplitude of the l th deflected beam of frequency $\omega + l\Omega$. The zero-order or main beam makes an angle θ_0 with the y axis in the x - y plane. The angle of deflection of the l th wave will be discussed shortly. Substituting (3) into (2) yields,

after performing the required algebra, a set of equations for the amplitudes V_l given by

$$\begin{aligned} d^2V_l/dy^2 - 2ik(\cos \theta_0)dV_l/dy + 2k(\cos \theta_0)\beta_l V_l \\ = -\frac{1}{2}k^2(1 + k\Omega/\omega)[([\Delta\epsilon/\epsilon]_c + i[\Delta\epsilon/\epsilon]_s)V_{l+1} \\ + ([\Delta\epsilon/\epsilon]_c - i[\Delta\epsilon/\epsilon]_s)V_{l-1}] \end{aligned} \quad (4)$$

in which

$$\beta_l = [2lkK(v/c' - \sin \theta_0) - l^2K^2(1 - v^2/c'^2)]/2k \cos \theta_0. \quad (5)$$

In the limit $\Delta\epsilon/\epsilon \ll 1$, V_l is a relatively slowly varying function of y and d^2V_l/dy^2 is negligible compared to the other terms in (4). Neglecting $\Omega/\omega \ll 1$ and defining

$$\xi(y) = \frac{1}{2}(k/\cos \theta_0)([\Delta\epsilon(y)/\epsilon]_c + i[\Delta\epsilon(y)/\epsilon]_s) \quad (6)$$

(4) can be rewritten (using $*$ to denote complex conjugate)

$$dV_l/dy + i\beta_l V_l = -\frac{1}{2}i[\xi V_{l+1} + \xi^* V_{l-1}]. \quad (7)$$

The initial conditions are $V_l(0) = 0$ for $l > 0$, $V_0(0) = V_0$. The equation for the deflected wave amplitudes, (7), has been found by many authors for the case $\xi = \text{constant}$ and real. Solutions have been found and are described in Ref. 1. More recent unpublished investigations⁴ have yielded the same results. In what follows the major emphasis will be given to the case $\xi = \xi(y)$ and to displaying the solution in a form which has a simple physical interpretation. Consistent with the experiments to be described, it is assumed that

$$\left| \int_{-\infty}^{+\infty} \xi(y) dy \right| \ll 1$$

and it follows that $V_{l+1} \ll V_l$. Thus to a very good approximation, for $l > 0$, (7) can be integrated to yield

$$V_l(y) \approx (\exp -i\beta_l y) \int_{-\infty}^y dy' [-\frac{1}{2}i\xi(y')^* V_{l-1}(y')] \exp +i\beta_l y'. \quad (8)$$

A similar equation holds for $l < 0$ with $\xi^* V_{l-1}$ replaced by ξV_{l+1} . Inspection of (8) indicates that all the V_l will be essentially zero unless $\beta_1 L < \pi$ (in which L is a measure of the interaction length or width of the acoustic beam). An exception may hold when ξ is a rapidly varying periodic function of y . This exception is of no interest here, since it corresponds to a situation where the acoustic beam has components moving at large angles relative to the x axis. When $\beta_1 L$ is small, the rest

of the $\beta_1 L$ cannot be small unless $K^2 L/k < \pi$. When this inequality holds, many grating orders can be excited and the interaction is said to be in the Raman-Nath regime. When $K^2 L/k > \pi$ the interaction is said to be in the Bragg scattering regime and only V_1 (or V_{-1}) and V_0 can be nonzero.

In the region in which the deflected beam is observed, that is, beyond the optical-acoustic interaction region, $\xi(y) \equiv 0$ and one can without error extend the upper limit of the integration in (8) to $+\infty$. Thus the amplitude of the deflected beam can be written

$$V_1 \approx (-\frac{1}{2}iV_0 \exp -i\beta_1 y) \int_{-\infty}^{+\infty} dy' \xi(y')^* \exp +i\beta_1 y'. \quad (9)$$

A similar equation holds for V_{-1} . It is assumed that because

$$\left| \int_{-\infty}^{+\infty} \xi(y) dy \right| \ll 1,$$

V_0 is constant. In fact V_0 differs from its initial value by terms of order

$$\left| \int_{-\infty}^{+\infty} \xi(y) dy \right|^2$$

and higher. As an example of the error produced by this approximation, consider the case where $\xi(y)$ is constant over a length L and zero elsewhere. Equation (9) predicts

$$V_1 \sim \sin \frac{1}{2}\beta_1 L / \beta_1$$

while (7) predicts, in the Bragg scattering limit,¹

$$V_1 \sim \sin \frac{1}{2}(\beta_1^2 + \xi^2)^{1/2} L / (\beta_1^2 + \xi^2)^{1/2}$$

which differs insignificantly when $\xi L \ll 1$. Note that V_1 becomes small when $\beta_1 L > \pi$, as noted earlier.

The quantity β_1 can be evaluated as a function of θ_0 from (5), which yields

$$\begin{aligned} \beta_{\pm 1} &= [\pm 2kK(v/c' - \sin \theta_0) - K^2(1 - v^2/c'^2)] / 2k \cos \theta_0 \\ &= \pm K(\sin \Theta - \sin \theta_0) / \cos \theta_0 \end{aligned} \quad (10)$$

in which Θ is an angle defined by

$$\sin \Theta = v/c' \mp \frac{1}{2}(K/k)(1 - v^2/c'^2). \quad (11)$$

The upper sign corresponds to $l = +1$, while the lower sign is for $l = -1$. If v/c' were zero, Θ would exactly equal the Bragg angle for scattering off the acoustic wavefront. In actuality the scattering plane

is rotated from the acoustic wavefront by an angle very closely equal to v/c' , which for acoustic waves can be neglected. In recognition of this fact Θ will be referred to as the Bragg angle.

Substituting (10) into (9) yields

$$V_1(\theta_0) \approx \left(-\frac{1}{2}iV_0 \exp + iKy(\sin \theta_0 - \sin \Theta)/\cos \theta_0\right) \\ \times \int_{-\infty}^{+\infty} dy' \xi(y')^* \exp -iKy'(\sin \theta_0 - \sin \Theta)/\cos \theta_0. \quad (12)$$

As stated earlier, the amplitude of the deflected wave, resulting from an interaction at angle θ_0 , is proportional to the Fourier transform of the acoustic wave amplitude. The relative intensity of the deflected light beam is given by $|V_1/V_0|^2$. Since $\xi(y)$ varies as $(\cos \theta_0)^{-1}$ [see (6)], the substitution $y'' = y'/\cos \theta_0$ puts (12) into a form which can be recognized as the expression for far-field or Fraunhofer diffraction for waves with propagation constant K . Thus $|V_1(\theta_0)|^2$ determines the far-field diffraction pattern of the acoustic beam.

The integral in (12) may also be interpreted by noting that an acoustic plane wave moving at a small angle ψ with respect to the x axis in the x - y plane can be described by a variation

$$\cos [\Omega t - Kx \cos \psi + Ky \sin \psi] = \cos [Ky \sin \psi] \\ \cdot \cos [\Omega t - Kx \cos \psi] - \sin [Ky \sin \psi] \sin [\Omega t - Kx \cos \psi].$$

From (1) and (6) it follows that

$$\xi(y) = \text{constant} \times \exp -iKy \sin \psi. \quad (13)$$

Inserting this value of ξ into the integral in (12) indicates that the deflected intensity is nonvanishing only for

$$\tan \psi = (\sin \theta_0 - \sin \Theta)/\cos \theta_0$$

or, since ψ is small, for $\psi \approx \theta_0 - \Theta$. Hence the deflected light intensity, when the main beam moves at angle θ_0 , measures only the component of the acoustic beam moving at angle $\psi \approx \theta_0 - \Theta$ relative to the x axis. By studying the variation of the deflected light intensity as a function of the angle θ_0 one determines the angular distribution of the acoustic energy. Since the diffraction angle of the light can, in practice, be kept small ($<10^{-4}$ radians) the angular resolution can be quite adequate except for the case of very high-frequency ($>10^{10}$ cps) acoustic waves.

From the above discussion it follows that for acoustic beams of finite width θ_0 will never differ much from Θ and (12) can be written

$$V_1(\theta_0) \approx [-\frac{1}{2}iV_0 \exp + iK(\theta_0 - \Theta)y] \times \int_{-\infty}^{+\infty} dy' \xi(y')^* \exp -iK(\theta_0 - \Theta)y'. \quad (14)$$

The equation for V_{-1} is the negative complex conjugate of (14) with the appropriate value of Θ for $l = -1$.

The angle of the deflected beam can be determined directly from (3). The l th deflected beam appears at an angle θ_l relative to the y -axis in the x - y plane defined by

$$(1 + \Omega/\omega) \sin \theta_l = \sin \theta_0 + lK/k. \quad (15)$$

In what follows $\Omega/\omega \ll 1$ will be neglected and only the case $l = \pm 1$ will be considered. Thus (15) can be rewritten [using (11)] as

$$\sin \theta_l = \sin \theta_0 - 2 \sin \Theta. \quad (16)$$

Defining the deflected angle as $\theta_l - \theta_0$, (16) can be solved to yield

$$\theta_l - \theta_0 \approx -2\Theta - (\Theta - \theta_0)^2 \tan \theta_0. \quad (17)$$

Terms of order $(\Theta - \theta_0)^4$ and higher have been neglected. The deflection angle has a magnitude very closely equal to $2|\Theta|$ and the variation in $\theta_l - \theta_0$ is quite small when θ_0 is varied through a small angle about Θ . This fact will prove to afford considerable experimental convenience.

When discussing the experimental results two distributions of acoustic intensity will be of interest. The angular dependence of the deflected intensity will be reviewed here to provide continuity in the discussion of the experiments.

Case I: A single acoustic beam of rectangular cross-section (see Fig. 2a).

For this case the acoustic beam of width L is assumed to have a constant amplitude

$$\begin{aligned} \xi(y) &= \xi & -\frac{1}{2}L \leq y \leq \frac{1}{2}L \\ &= 0 & \text{elsewhere} \end{aligned}$$

and

$$V_1(\theta_0) = (-\frac{1}{2}i\xi^*V_0 \exp + iK(\theta_0 - \Theta)y)$$

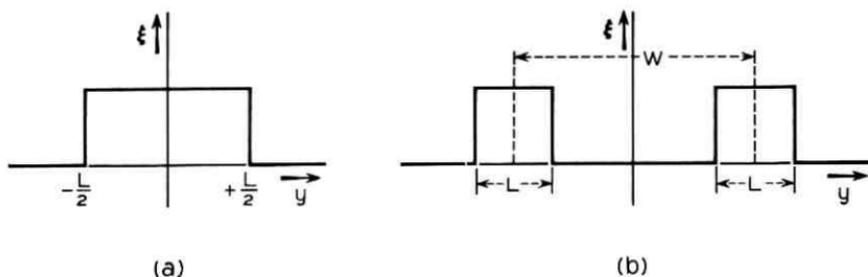


Fig. 2—Schematic representation of acoustic beam cross-sections; (a) single beam, (b) parallel coherent beams.

$$\begin{aligned} & \int_{-L/2}^{+L/2} dy' \exp -iK(\theta_0 - \Theta)y' \\ & = \left(-\frac{1}{2}i\xi^*LV_0 \exp +iK(\theta_0 - \Theta)y \right) \left[\frac{\sin \frac{1}{2}K(\theta_0 - \Theta)L}{\frac{1}{2}K(\theta_0 - \Theta)L} \right]. \end{aligned} \quad (18)$$

The angular dependence is precisely the same as that for single-slit Fraunhofer diffraction. Note that the first zeros on either side of the central maximum are separated by an angle

$$\Delta\theta_0 = 4\pi/KL \quad (19)$$

which is a direct measure of the acoustic beamwidth L .

Case II: Two parallel coherent acoustic beams with phase difference ϑ (see Fig. 2b).

From (6) it follows that

$$\xi(y) \cos (\Omega t - Kx - \varphi) \rightarrow \xi(y) (\exp i\varphi) \cos (\Omega t - Kx) \dagger$$

hence

$$\begin{aligned} \xi(y) &= \xi & -\frac{1}{2}(W+L) \leq y \leq -\frac{1}{2}(W-L) \\ &= \xi \exp i\varphi & \frac{1}{2}(W-L) \leq y \leq \frac{1}{2}(W+L) \\ &= 0 & \text{elsewhere.} \end{aligned}$$

Each beam has width L and the center-to-center spacing is W . For this case

$$\begin{aligned} V_1(\theta_0) &= \left(-\frac{1}{2}i\xi^*V_0 \exp iK(\theta_0 - \Theta)y \right) \\ & \cdot \left[\int_{-1/2(W+L)}^{-1/2(W-L)} dy' \exp -iK(\theta_0 - \Theta)y' \right] \end{aligned}$$

† This is not an equality but is written to indicate that a change in the phase of the acoustic wave implies an equivalent change in the argument of $\xi(y)$.

$$\begin{aligned}
& + (\exp -i\varphi) \int_{\frac{1}{2}(w-L)}^{\frac{1}{2}(w+L)} dy' \exp -iK(\theta_0 - \Theta)y' \quad (20) \\
& = (-i\xi^*LV_0 \exp i[K(\theta_0 - \Theta)y - \frac{1}{2}\varphi]) \left[\frac{\sin \frac{1}{2}K(\theta_0 - \Theta)L}{\frac{1}{2}K(\theta_0 - \Theta)L} \right. \\
& \quad \left. \times \cos \frac{1}{2}[K(\theta_0 - \Theta)W + \varphi] \right].
\end{aligned}$$

As would be expected, the angular dependence is precisely the same as for double-slit Fraunhofer diffraction (Young's experiment).

One other point is perhaps worth noting in passing. The relative intensity of the deflected light beam at the optimum angle is given by

$$|\frac{1}{2}\xi L|^2 = \frac{1}{16}[(\Delta\epsilon/\epsilon)_c^2 + (\Delta\epsilon/\epsilon)_s^2]k^2L^2/\cos^2\theta_0$$

which is directly proportional to the acoustic power. Quantitative determination of the acoustic power in the transmission medium can be made if the photoelastic and elastic constants of the material are known. Thus direct measurements of transducer efficiency can be made.

1.2 Experiment

The experiments to be described were performed using ultrasonic waves in the frequency range 50–250 mc/s. The transmission medium was fused quartz ($v = 5.96 \times 10^5$ cm/s) with rectangular cross-section of $\frac{1}{4} \times \frac{1}{2}$ inch and length one inch. A thin-film cadmium sulphide longitudinal wave transducer was evaporated onto one end of the delay line.† The faces through which the light entered and left the medium were optically flat and parallel and were antireflection coated for 6328-Å radiation. In some cases the face opposite the transducer was terminated with a mercury pool, which reduced the reflected amplitude by about 10 db.

Light from a 6328-Å He-Ne laser operating in the lowest-order transverse mode with a power of a few milliwatts was polarized at 45° to the direction of acoustic propagation. The acoustically produced strain makes the quartz uniaxial; for longitudinal acoustic waves the optic axis is along the direction of propagation. Hence, for the geometry employed, the privileged axes corresponded closely to the x and z axes and $\Delta\epsilon$ for E_z differed from $\Delta\epsilon$ for E_x . As a result the deflected beam polarization was rotated from the 45° position. This was convenient experimentally, since the zero-order or main beam could be effectively eliminated with a crossed analyser without drastically attenuating the

† The transducers were prepared by N. F. Foster.

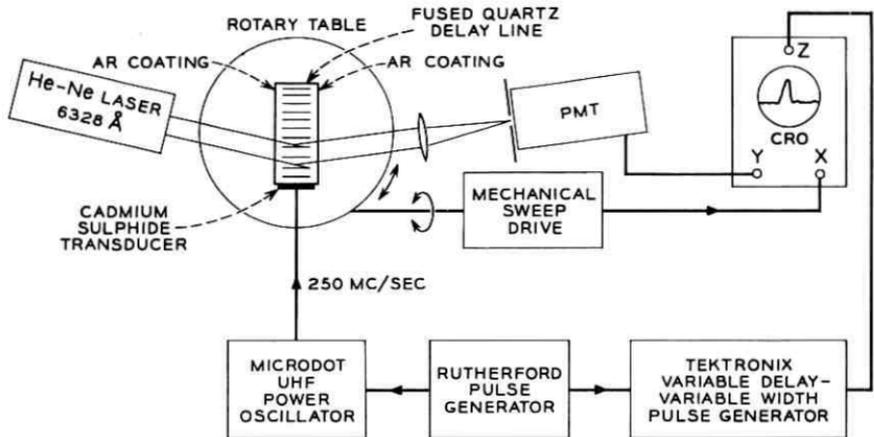


Fig. 3 — Experimental arrangement for observing angular dependence of deflected light intensity.

deflected beam. A lens focused the deflected beam through a narrow slit mounted in front of a photomultiplier. Deflection angles were of the order of one degree.

The experimental arrangement is shown in Fig. 3. When necessary the acoustic energy was pulse modulated. A mechanical table rotator was also used since, as noted earlier, the angular position of the deflected beam was essentially independent of the orientation of the acoustic wave relative to the incident light beam, and it was unnecessary to change the photomultiplier or laser position as θ_0 was varied. Thus by supplying a voltage proportional to the angular rotation of the table to the X axis of the oscilloscope and the photomultiplier output to the Y axis, the deflected energy as a function of θ_0 could be displayed. A baseline was produced by pulse modulation of the acoustic energy.

A typical display is shown in Fig. 4 and very closely follows the predicted $[\sin \frac{1}{2}K(\theta_0 - \Theta)L / \frac{1}{2}K(\theta_0 - \Theta)L]^2$ behavior for a single acoustic beam of rectangular cross-section. It should be noted that the angle θ_0 is that of the light within the medium and is smaller by the factor $n = 1.46$ than the angle measured on the table because of refraction in the quartz. (Severe errors are introduced when the surfaces are not optically antireflection coated, because of multiple reflection. The multiple reflection can be used to greatly enhance the optical-acoustic interaction and will be discussed in more detail in a forthcoming paper.) Using (19), the measured acoustic beamwidth corresponded very closely to the width of the transducer. For very wide transducers (> 2 cm)

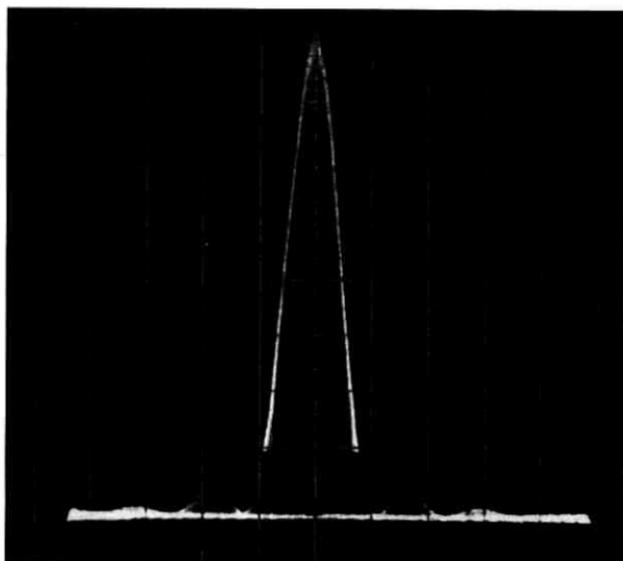


Fig. 4—The angular dependence for a rectangular beam cross-section which closely followed the theoretical $(\sin x/x)^2$ behavior. The angular separation between zeros adjacent to the central peak was 0.6° at the frequency of 248 mc; using this value in (19) one calculates the width of the acoustic beam to be 4.2 mm, which corresponds very closely to the width of the transducer. The peak deflected light intensity was in the range 0.1 to 1 per cent of the intensity of the zero-order beam. The acoustic intensity was of the order of 0.1 watt/cm².

deviations from the ideal $(\sin x/x)^2$ behavior were observed and ascribed to nonuniform transducers and finite resolution associated with diffraction of the light. The width of the beam was still described by (19), however.

The angle for optimum interaction, $\theta_0 = \Theta$, is predicted by (11). The difference between Θ for deflection into the $l = +1$ beam (Θ_+) and Θ for deflection into the $l = -1$ beam (Θ_-) is given by

$$\begin{aligned} \Theta_+ - \Theta_- &\approx \sin \Theta_+ - \sin \Theta_- \\ &= -K/k = -(\Omega/\omega)(c'/v) \end{aligned} \quad (21)$$

using (11) with $v^2/c'^2 \ll 1$. Direct verification of (21) in the range $\Omega/2\pi = 50$ –250 mc/s was obtained by using a narrow slit in front of the photomultiplier and determining the angular difference between the peaks of the $(\sin x/x)^2$ curves for $l = \pm 1$.

The angle of deflection is given as $\theta_l - \theta_0 \approx (lK/k)$. This was verified also for $l = \pm 1$. No higher orders were observed, consistent with the

fact that $K^2L/k > \pi$, placing the interaction in the Bragg scattering regime.

It is interesting to note that when the delay line is unterminated an acoustic standing wave is established. The $l = +1$ deflected wave at frequency $\omega + \Omega$ associated with one of the traveling acoustic waves, and the $l = -1$ wave at frequency $\omega - \Omega$ associated with the oppositely moving acoustic wave are deflected into the same angle [see (16)] and can mix in the photodetector. A beat at the difference frequency $2(\Omega/2\pi)$ was detected, and it was found as expected that the light was essentially 100 per cent modulated. The amplitude of the beat as a function of θ_0 also followed the $(\sin x/x)^2$ variation.

Fig. 5 illustrates the results of an experiment in which the frequency of the acoustic wave was swept periodically. When the deflected beam was allowed to travel several feet and fall on the wall of the room the reciprocating motion of the beam could easily be observed. The oscilloscope display shows the output of the photomultiplier as a function of the acoustic frequency. The horizontal axis is approximately linear with acoustic frequency. The narrow slit in front of the photomultiplier sampled the Gaussian light distribution in the deflected beam as it was swept across the slit. The "frequency width" of the beam using (16) agreed well with the "angular width" of the beam associated with diffraction. Successive photographs (top to bottom) show decreasing sweep rates and sweep widths. When the sweep rate was sufficiently slow, the acoustic resonances, separated by approximately 120 kc/s, built up the acoustic energy to its steady-state value as seen in the bottom photograph. The unequal spacing of the acoustic resonances can be attributed to some residual FM in the swept oscillator which produced a slightly nonuniform frequency base. The vertical gain was kept constant, and it can be seen that no great advantage was obtained in the resonant condition because the acoustic Q was not very high. The attenuation of the fused quartz was measured to be 0.6 db/cm or about 4.5 db for the round trip.

An interesting property of the transducer was uncovered during the course of the experiments. The far end of the delay line was unterminated and the driving signal was pulse modulated with pulses short compared to the length of the delay line but long compared to the width of the optical beam. The spacing between applied pulses was sufficiently long that there was no overlap between each successive decaying pulse train. Synchronized, variable-delay pulses were applied to the Z axis of the oscilloscope with the correct timing so that the optical-acoustic interaction was observed for only one of the pulses in the

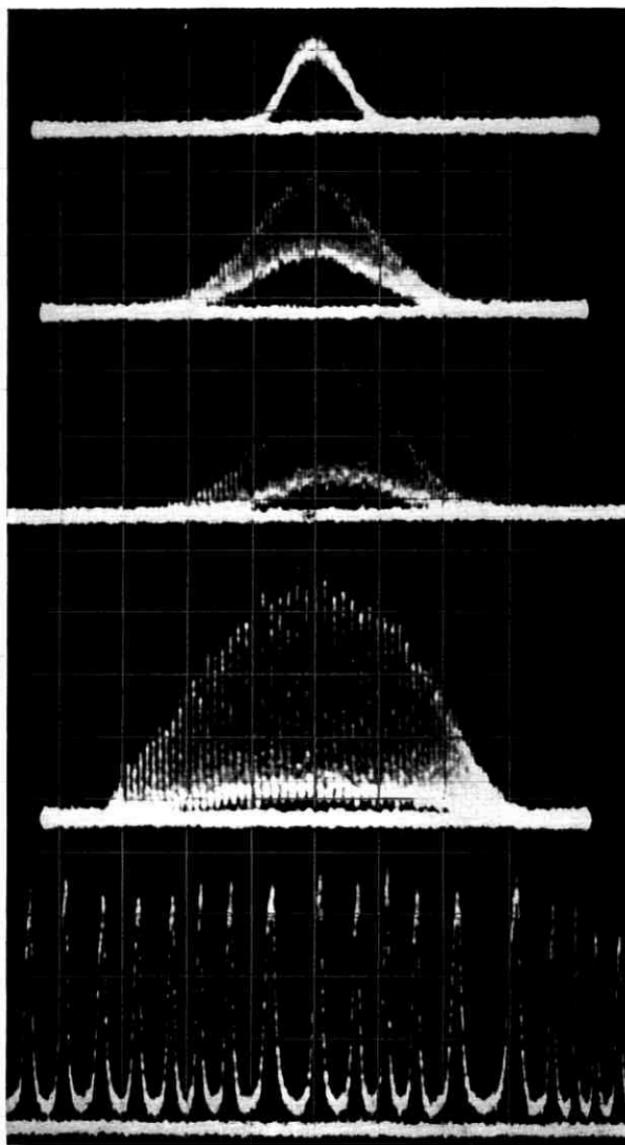


Fig. 5—Light intensity at a fixed angular position as a function of acoustic frequency. The photomultiplier samples the intensity of the approximately Gaussian beam as it is swept across the defining slit.

decaying train. Successive pulses in the train could be studied by increasing the delay.

Fig. 6 illustrates the appearance of the $(\sin x/x)^2$ display associated with each of the pulses in the train. Multiple exposures were taken to

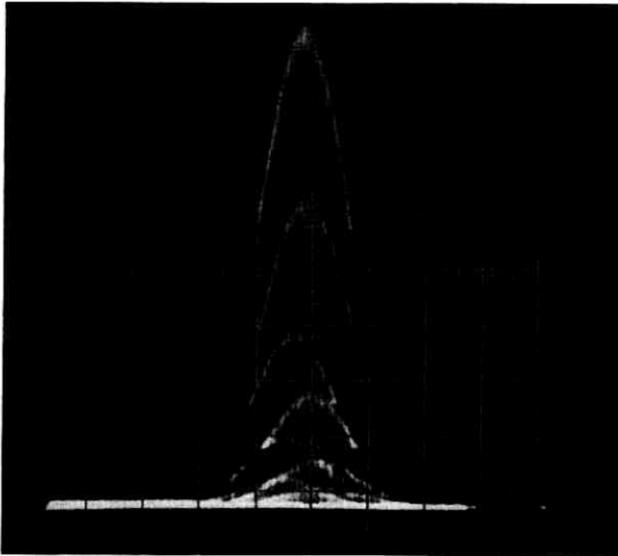


Fig. 6 — Multibounce appearance of optical-acoustic interaction indicating angular divergence of successively reflected beams. The angular divergence is less than 0.1° .

record the data on one photograph. In addition, each exposure covered about 20–50 angular sweeps and each angular sweep covered many thousands of pulses. Thus there is no doubt that the data were entirely reproducible and that the observed displacement of the optimum angle for successive pulses in the train was real.

The photograph can be interpreted in the following way. The face of the delay line opposite the transducer was optically flat and parallel to the transducer face. Thus, if the acoustic wave were launched in a direction normal to the face, it would be expected to follow the same path after reflection from the opposite face at normal incidence (see Fig. 7). For the first passage of the forward pulse, the peak of the $(\sin x/x)^2$ curve indicates the optimum angle for the $l = 1$ interaction and, incidentally, marks the precise direction of the first forward wave. If the first backward pulse were exactly parallel to the first forward pulse, the peak of the $l = -1$ interaction should have appeared at precisely the same angle, although reduced in amplitude because of the acoustic loss. This is seen to be the case. However, the second forward pulse was slightly displaced in angle from the first forward pulse, indicating that it was not parallel to the first forward pulse. Since the opposite face should reflect specularly, the second backward pulse

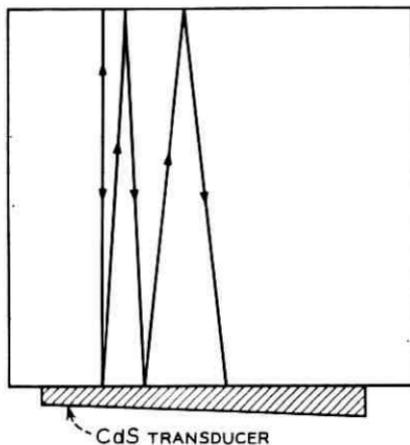


Fig. 7 — Postulated paths of successively reflected beams. The angular spread is greatly exaggerated.

should have been displaced by the same angle from the first backward pulse except in the opposite direction. This is also seen to be the case. The third forward pulse was displaced again from the second forward pulse by the same amount and in the same direction as the second was displaced from the first.

The behavior of these and succeeding pulses (see Fig. 7) can be explained by assuming that reflection of the transducer face was specular with respect to a surface which was not parallel to the opposite face. This could happen if the transducer were wedge-shaped but somehow launched an acoustic wave normal to the delay line interface. Such behavior could introduce spurious results in conventional pulse-echo experiments, since the angular dependence of the transducer response is given essentially by the same $(\sin x/x)^2$.

Further corroboration of the angular variation appeared when the delay line was allowed to resonate. The acoustic intensity in the beam cross-section showed variations which did not appear in the pulsed beam.

A more critical test of the Fourier transform relation between the acoustic intensity and the angular dependence of the deflected beam was performed using the arrangement shown in Fig. 8. In this case two identical terminated delay lines were separated by a fused quartz parallel flat. The 250 mc/s signal was divided, isolated, and applied to each transducer with a variable phase difference. The characteristic $(\sin x/x)^2$ behavior for each delay line individually peaked at the same

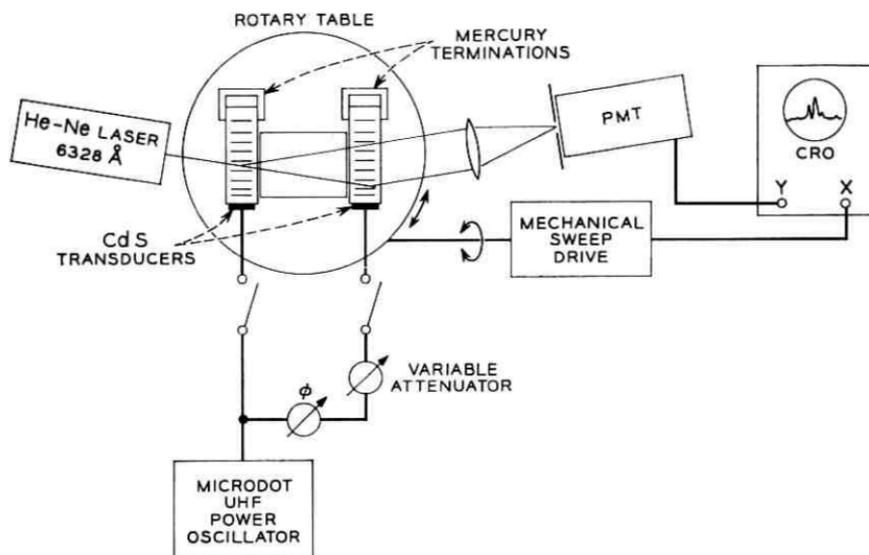


Fig. 8—Experimental arrangement for double acoustic beam interaction.

angle and with the same amplitude, indicating that the acoustic beams were aligned and had the same intensity.

The photograph in Fig. 9 exhibits the behavior predicted by (20). The envelope has the behavior $[\sin \frac{1}{2}K(\theta_0 - \Theta)L / \frac{1}{2}K(\theta_0 - \Theta)L]^2$ superimposed on the $\cos^2 \frac{1}{2}[K(\theta_0 - \Theta)W + \varphi]$ behavior required by the spacing W . The lack of a perfect zero can be ascribed to the finite angular spread of the light beam, which was not negligibly small compared to the spacing between minima, $2\pi/KW$, and thereby tended to slightly wash out the perfect $\cos^2 x$ behavior. The observed number and spacing of the minima are consistent with the spacing W . In Fig. 10 the upper photograph was taken for $\varphi = 0$ and the lower photograph for $\varphi = \pi/2$. The peak deflected intensity is four times that for each beam, in agreement with (20).

II. FINITE VOLUME LOSS

2.1 Theory

When the acoustic beam is attenuated, (1), which describes the acoustic wave propagation, must be modified by including a factor $\exp -\alpha x$, in which 2α is the reciprocal decay distance for the acoustic energy. If the intensity of the acoustic wave is sufficiently low that the conditions of Section I

$$\left(\left| \int \xi dy \right| \ll 1 \right)$$

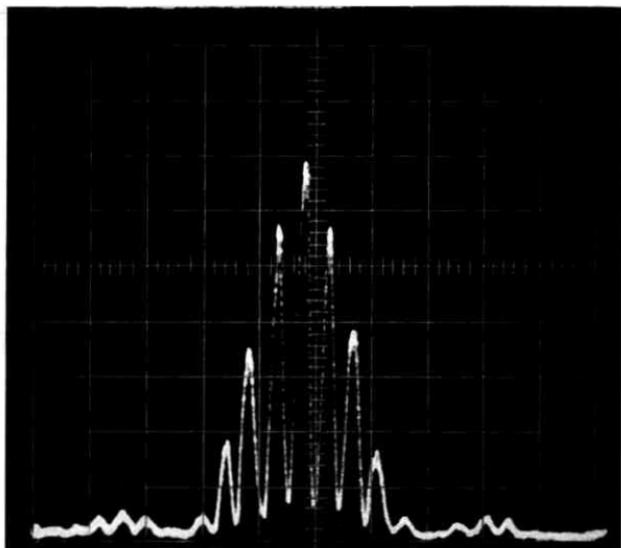


Fig. 9—The angular distribution for two coherent acoustic beams.

are satisfied, then a suitable solution for the deflected light amplitude can be written

$$\begin{aligned}
 E(x, y, t) = \sum_{l=-\infty}^{+\infty} V_l(y) \exp -|l|\alpha x \exp i[(\omega + l\Omega)t \\
 - (k \sin \theta_0 + lK)x \\
 - ky \cos \theta_0] + \text{complex conjugate.}
 \end{aligned}
 \tag{22}$$

Note that (22) differs from (3), the solution for the nondecaying acoustic beam, by the factor $\exp -|l|\alpha x$ associated with the l th deflected beam. Substituting into (2) yields an equation identical to (4) except that the term in V_{l+1} (for $l > 0$) has the factor $\exp -2\alpha x$. Therefore, strictly speaking, (22) is not a proper solution. However, when $V_{l+1} \ll V_{l-1}$, to a very good approximation, (22) represents the scattered wave amplitudes and V_l is defined by (4). One additional difference arises: the term β_l is no longer given by (5) but rather must be written

$$\beta_l' = \beta_l + il\alpha(\sin \theta_0 + lK/k)/\cos \theta_0
 \tag{23}$$

in which β_l is given by (5). A term in $(\alpha/K)^2 \ll 1$ has been left out. The significant point here is that β_l' can never be zero even when $\theta_0 = \Theta$, corresponding to the optimum deflection condition. In this

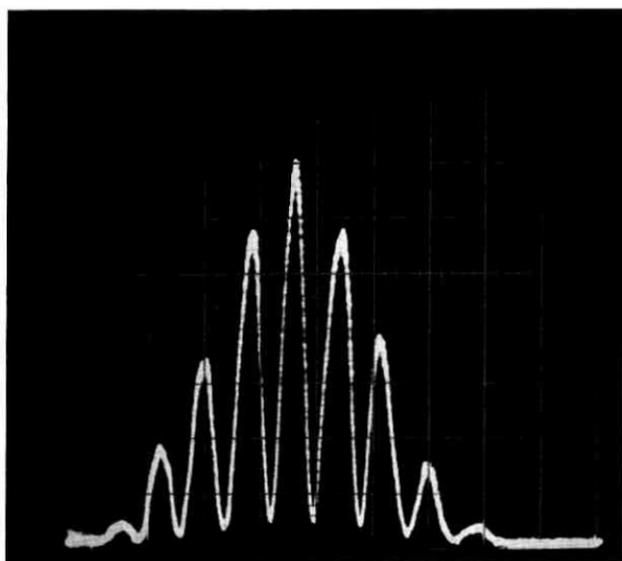
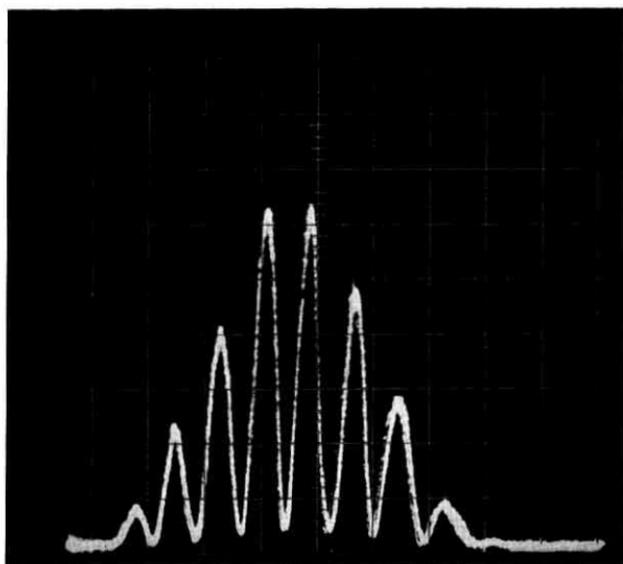
 $\phi = 0$  $\phi = \pi/2$

Fig. 10—The angular distribution for $\phi = 0$ (upper) and $\phi = \pi/2$ (lower).

region, which is the only region of interest, $K/k \approx -2 \sin \theta_0$ and β_1' can be written

$$\beta_1' \approx \beta_1 - i\alpha \tan \theta_0. \quad (24)$$

With the substitution of β_1' for β_1 all the results of Section I are valid for the angular dependence of the optical-acoustic interaction. Thus

$$-K(\theta_0 - \Theta) \rightarrow -K(\theta_0 - \Theta) - i\alpha \tan \theta_0. \quad (25)$$

As an example of the consequences of this substitution, consider the angular dependence of the scattered intensity $|E_1(\theta_0, x)|^2$ corresponding to $l = 1$. For finite α , (18), derived for a rectangular distribution and an infinitely wide light beam, becomes

$$|E_1(\theta_0, x)|^2 = [\exp -2\alpha x - \alpha L \tan \theta_0] \frac{|\sin \frac{1}{2}[K(\theta_0 - \Theta) + i\alpha \tan \theta_0]L|^2}{\frac{1}{4}[K^2(\theta_0 - \Theta)^2 + \alpha^2 \tan^2 \theta_0]L^2}. \quad (26)$$

Nonessential constants have been suppressed and $|E_1|^2$ normalized so that it has the value unity for $\theta_0 = \Theta$, $\alpha = 0$. This result has a straightforward interpretation. Since the scattered beam has an exponential decay in the cross section, the effective beamwidth is of order $(2\alpha)^{-1} \cos \theta_0$ and the angular spread is $2\alpha/k \cos \theta_0$.† The product $K(2\alpha/k \cos \theta_0)$ gives rise to the term $\alpha \tan \theta_0$ in (26) and limits the angular resolution in observing the variation of $|V_1(\theta_0)|^2$.

For scattering at the Bragg angle, $\theta_0 = \Theta$,

$$|E_1(\Theta, x)|^2 = \frac{[\exp -2\alpha x - \alpha L \tan \Theta] \sinh^2 \frac{1}{2}\alpha L \tan \Theta}{\frac{1}{4}\alpha^2 L^2 \tan^2 \Theta}. \quad (27)$$

It is instructive to note that this result can be obtained from a very simple ray picture of the scattering interaction. Consider Fig. 11, which shows a typical scattered ray for interaction at the Bragg angle. Along this path, defined by $x = x_0 + y \tan \Theta$, the contributions from each scattering point are additive and proportional to the product of the acoustic and light amplitude. Thus one can write for the normalized scattered intensity

$$|E_1(\Theta, x_0)|^2 = \left| L^{-1} \int_0^L dy \exp -\alpha(x_0 + y \tan \Theta) \right|^2. \quad (28)$$

Evaluating the integral yields a result identical to that given by (27).

† See (36).

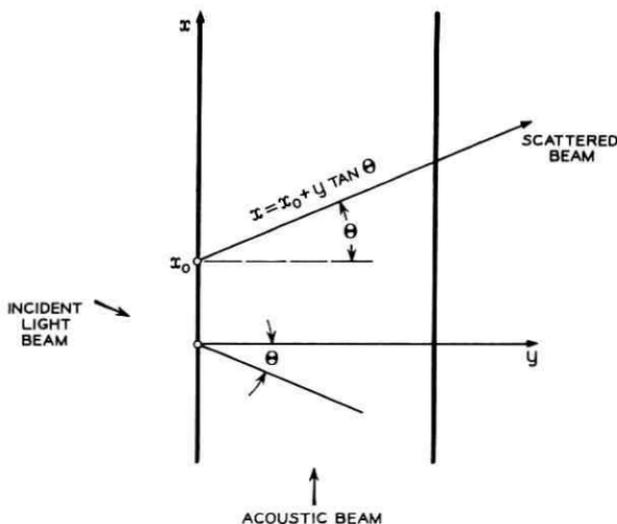


Fig. 11 — A typical scattered ray for interaction at the Bragg angle.

Note that the intensity of the scattered light is reduced from its value at $\alpha L = 0$ and can be related directly to the decreasing amplitude of the acoustic wave along the ray path. This effect will be referred to as "finite coherence width," which arises because of the acoustic decay and defines a maximum useful acoustic beam width $L = (\alpha \tan \Theta)^{-1}$. Note that the coherence width is important relative to the considerations of Section I only for $\alpha \approx K$.

In what follows the relatively simple geometrical picture described above will be exploited to determine the near-field distribution of the scattered light from incident light beams with a Gaussian intensity distribution as well as for uniform beams of finite width. From this it will be possible to calculate the shape of the far-field diffraction pattern of the scattered light. Experimental confirmation of these far-field distribution patterns, which will be described later, constitutes a much more rigorous test of the theory than direct observation of the near-field pattern. In addition, it often offers considerable experimental convenience.

2.2 Diffraction of the Scattered Light

The results of the previous section indicate that when the acoustic wave amplitude decays as $\exp -\alpha x$ then the scattered beam will have superposed on its normal spatial dependence the factor $\exp -\alpha x$. Thus,

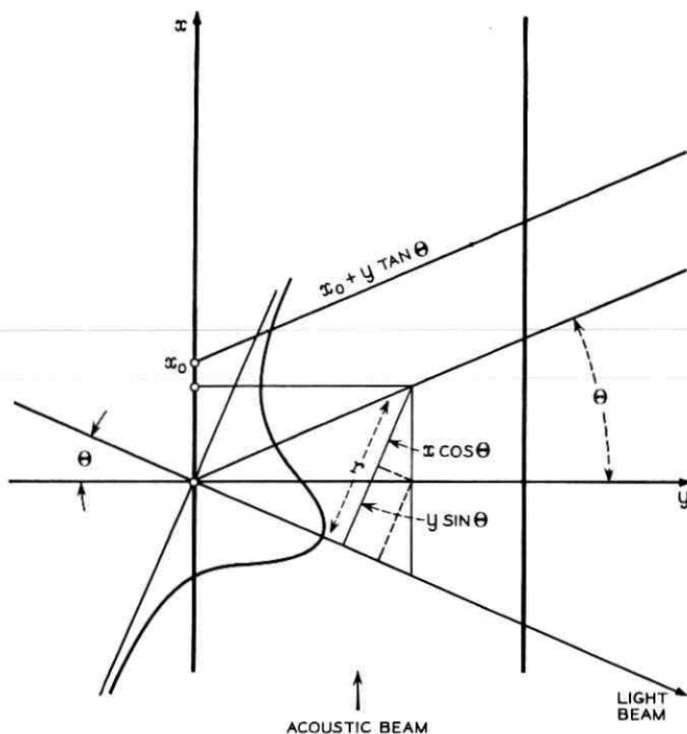


Fig. 12—Beam with Gaussian amplitude distribution incident at the Bragg angle.

if the light beam is translated in the x direction by an amount Δx , the scattered light energy will be changed by an amount $\exp -2\alpha\Delta x$ independent of the width of the light beam. Thus it is not necessary to use narrow light beams to probe the exponential decay of the acoustic energy.

In cases where it is inconvenient or undesirable to translate the light beam relative to the acoustic beam, the decay constant can be determined by studying the near-field decay of the scattered light beam in a direction parallel to the acoustic beam. In performing this experiment, it is convenient to use a Gaussian beam as obtained directly from a gas laser (TEM₀₀ mode). However, the beam must be modified, as can be seen from the following discussion.

If the incident beam has a Gaussian distribution, so that

$$E_0(x,y) = \exp -\frac{1}{2} \left[\frac{x \cos \Theta + y \sin \Theta}{w} \right]^2 \quad (29)$$

as shown in Fig. 12, in which w is related to the spot size of the beam,

then it is shown in Appendix A that the scattered beam amplitude can be written

$$E_1(\Theta, x_0) = \exp -(\alpha x_0 + \frac{1}{2}(x_0/w)^2 \cos^2 \Theta) \\ \times \left\{ X^{-1} \exp Y^2 \left[\int_0^{Y+X} \exp -y^2 dy \right. \right. \\ \left. \left. - \int_0^Y \exp -y^2 dy \right] \right\} \quad (30)$$

with

$$X = \sqrt{2} (L/w) \sin \Theta \\ Y = \frac{1}{4} \sqrt{2} (\alpha w / \cos \Theta + 2(x_0/w) \cos \Theta)$$

where the variable x_0 denotes that the scattering originates from a point $x = x_0, y = 0$ on the incident beam. When $X \ll 1$, the term in braces can be expressed as

$$1 - \frac{1}{2}[\alpha L \tan \Theta + (x_0/w)(L/w) \sin 2\Theta]. \quad (31)$$

The term $\alpha L \tan \Theta$ arises from the coherence width of the acoustic beam. In fact, for $w \rightarrow \infty$, (30) and (27) are identical. The linear term x_0 indicates that the distribution is no longer perfectly Gaussian nor is it symmetrical. Except for this slight deviation, however, the distribution is a displaced Gaussian with the same spot size as the incident beam. Thus the scattered beam shape contains essentially no information concerning the decay constant of the acoustic wave. This situation can be changed merely by partly blocking the incident Gaussian beam so that, for example, the part of the beam described by x less than some value x_0 is removed. Under this circumstance the shift in the peak of the Gaussian can be discerned and α can be determined. Alternately, the far-field diffraction of the scattered light beam can be observed. This is given by

$$V_1(\theta) = \int_{x_0}^{+\infty} dx E_1(\Theta, x) \exp -ik(\sin \theta - \sin \Theta)x. \quad (32)$$

Using $E_1(\Theta, x)$ as given in (30), neglecting the small change resulting from the term in braces, yields

$$V_1(\eta) = \sqrt{\pi/2} w^{-1} \cos \Theta [\exp \frac{1}{2} \alpha^2 w^2 (1 + i\eta)^2 / \cos^2 \Theta] \cdot \\ \operatorname{erfc} [(x_0/\sqrt{2} w) \cos \Theta + \alpha w (1 + i\eta) / \sqrt{2} \cos \Theta] \quad (33)$$

in which erfc denotes the complementary error function and

$$\eta = (k/\alpha)(\sin \theta - \sin \Theta). \quad (34)$$

For the unblocked beam ($x_0 = -\infty$), the scattered intensity has the form, exclusive of constants,

$$|V_1(\eta)|^2 \propto \exp \alpha^2 w^2 \eta^2 / \cos^2 \Theta = \exp k^2 w^2 (\sin \theta - \sin \Theta)^2 / \cos^2 \Theta$$

which, as would be expected, is Gaussian in angular distribution. In particular, the angular spread is independent of α . Contrast this behavior with the situation in which approximately half the Gaussian beam is blocked, so that $|x_0/w| \ll 1$ and the beam is wide compared to the acoustic decay distance $\alpha w \gg 1$. In this limit

$$\lim_{z \rightarrow \infty} (\text{erfc } z = (\sqrt{\pi} z)^{-1} \exp -z^2)$$

$$|V_1(\eta)|^2 \propto 1/(1 + \eta^2) \quad (35)$$

which is Lorentzian with an angular width at half power determined by (using 34)

$$\Delta\theta \approx 2\alpha/k \cos \Theta. \quad (36)$$

It is perhaps instructive to rewrite the parameter η in a slightly different form. Suppose an experiment is performed, similar to that associated with Fig. 5, in which the scattered light beam is observed at a fixed angle Θ but the acoustic frequency is varied. Clearly the angle θ can be considered to represent the peak of the energy distribution as a function of the acoustic frequency. Using (15), which defines the scattering angle for a given acoustic frequency and a given \mathbf{k} , η can be written

$$\eta = 2(\Omega - \Omega_0)\tau \equiv 2(\omega - \omega_0)\tau \quad (37)$$

in which Ω_0 is the acoustic frequency corresponding to interaction at angle Θ and Ω corresponds to angle θ . The parameter $\tau = (2\alpha v)^{-1}$ corresponds to the phonon lifetime if $(2\alpha)^{-1}$ is interpreted as the phonon mean-free path. As would be expected, the full width of the Lorentzian at half power is τ^{-1} . The difference $\Omega - \Omega_0$ also appears at the optical frequencies.

The basic conclusion here is that one can study phonon lifetimes (or mean-free paths or acoustic decay distances) by observation of the Lorentzian linewidth in a scattering experiment such as described, but one cannot use a full Gaussian beam such as might be obtained from a gas laser. Rather, a half-Gaussian beam whose width is large compared to the decay distance is required. A discussion of this result relative to the case of Brillouin scattering with thermally generated phonons is beyond the scope of this paper and is reserved for future consideration.

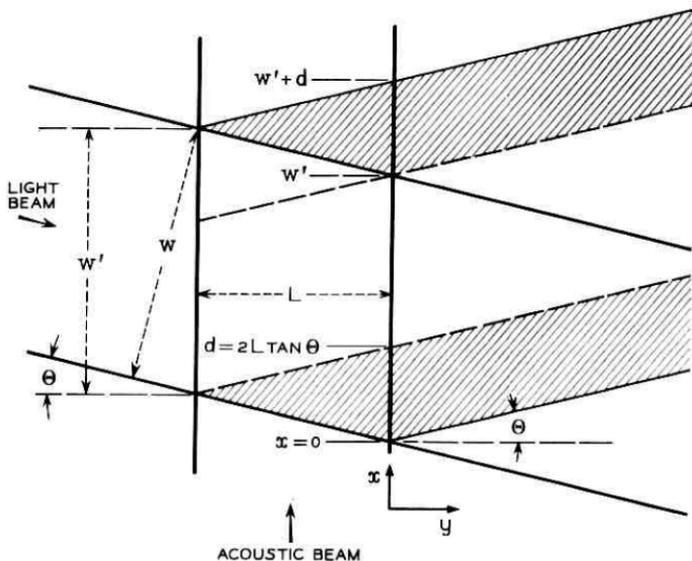


Fig. 13—Uniform beam of finite width w incident at the Bragg angle.

Another distribution of interest because of experimental simplicity is the uniform beam of width w . With reference to Fig. 13, the near-field distribution of the scattered beam amplitude for interaction at the Bragg angle may be written as

$$\begin{aligned}
 E_1(\Theta, x) &= (x/d) \exp -\alpha x & 0 \leq x < d \\
 &= \exp -\alpha x & d \leq x \leq w' \\
 &= [(w' + d - x)/d] \exp -\alpha x & w' < x \leq w' + d
 \end{aligned} \quad (38)$$

in which $d = 2L \tan \Theta$ and $w' = w/\cos \Theta$. The far-field distribution can be written as

$$V_1(\eta) = \int_0^{w'+d} dx E_1(\Theta, x) \exp -i\alpha\eta x \quad (39)$$

which yields after some manipulation

$$V_1(\eta) = \left[\frac{1 - \exp -\alpha(1 + i\eta)w'}{\alpha(1 + i\eta)} \right] \left[\frac{1 - \exp -\alpha(1 + i\eta)d}{\alpha(1 + i\eta)d} \right]. \quad (40)$$

When the coherence width is sufficiently large that $\alpha d = 2\alpha L \tan \Theta \ll 1$, the scattered energy distribution has the form

$$|V_1(\eta)|^2 = \frac{1 + \exp -2\alpha w' - 2[\exp -\alpha w'] \cos \alpha w' \eta}{\alpha^2(1 + \eta^2)}. \quad (41)$$

Note that the angular distribution is Lorentzian only in the limit $\alpha w' \gg 1$. When $\alpha w'$ is not large, diffraction associated with the beam of width w modifies the line shape. (Even when $\alpha = 0$, if d/w' is not very small, the "edge effect" will change the far-field diffraction pattern.)

When αd is large so that the coherence width dominates, the line shape becomes a Lorentzian squared. In the intermediate region where αd is small but not negligible, one can write (40) in the form

$$|V_1(\eta)|^2 \approx \frac{1 - \alpha d + (1/12)\alpha^2 d^2 [7 - \eta^2]}{1 + \eta^2}. \quad (42)$$

Thus the line shape is essentially Lorentzian except that the wings have a value somewhat less than that of a true Lorentzian.

2.3 Experiment

The experiments with finite α were performed using water at about 250 mc/s. The water cell consisted of two parallel quartz flats which were antireflection coated on the air side. Reflection at the quartz-water interface was negligible. One end of the water cell consisted of the fused quartz delay line described in Section I, which acted as a resonant buffer rod with evaporated-layer CdS as the longitudinal wave transducer. The opposite end of the cell was sufficiently far away that no reflections occurred. The traveling acoustic wave was square-wave modulated at 1 kc/s to allow synchronous detection of the scattered light.

A narrow slit was mounted in front of a photomultiplier and the entire assembly mounted on a micrometer-driven stage. The intensity distribution of the scattered light as well as that of the main beam could be mechanically scanned in the near or far field.

Near-field traces were taken with the slit as close to the water cell as possible; far-field traces were taken with the slit in the focal plane of a $\frac{1}{2}m$ focal length lens. The output of the phase-sensitive detector was applied to the Y axis of an X-Y recorder. The X axis was driven by the reference voltage from a Hewlett-Packard sweep drive unit attached to the micrometer stage. Visual confirmation of the recorder traces was made by observing the scattered light with a telescope focused on the cell for near field, and to infinity for the far field. The acoustically

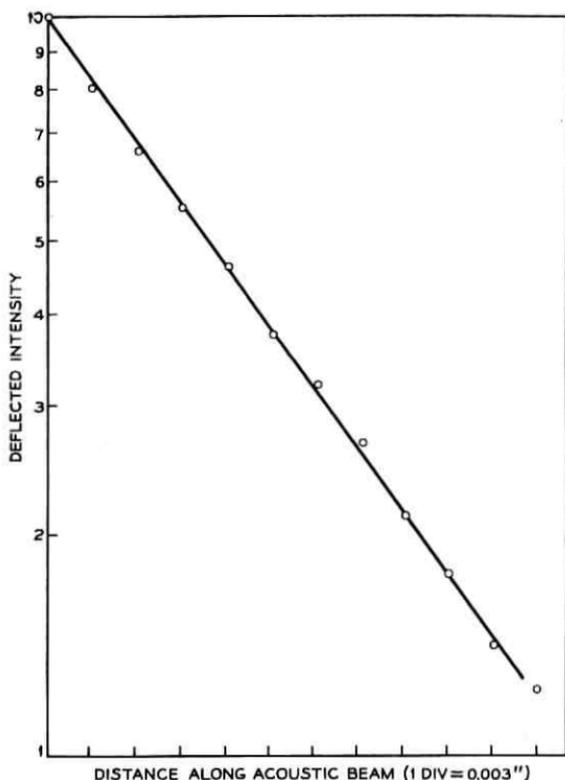


Fig. 14 — Deflected light intensity in arbitrary units vs distance along the acoustic beam.

scattered light could easily be distinguished from other “dirt” scattered light, since there was no granularity as is common for stationary scatterers.⁵ Acoustic streaming⁶ could also be observed.

As expected, the far-field diffraction of the acoustic beam, observed as described in Section I, exhibited little “coherence-width” degradation, i.e., the angular smear introduced by the term $\alpha \tan \theta_0$ in (26) was only 10^{-2} of the zero spacing. In what follows all observations were made at $\theta_0 = \Theta$.

The decay constant α was measured using a Gaussian beam from the laser and translating it along the acoustic beam. Fig. 14 is a semilog plot of the deflected intensity in arbitrary units vs distance. The curve yields the value $\alpha = 12.6 \text{ cm}^{-1}$. For comparison, it is necessary to extrapolate values of α measured at lower frequency.⁶ The measured value $\alpha/f^2 = 2.1 \times 10^{-16} \text{ (cm}^{-1} \text{ sec}^2)$ (f is the acoustic frequency = 245 mc/s) gives results in good agreement with the low-frequency results.

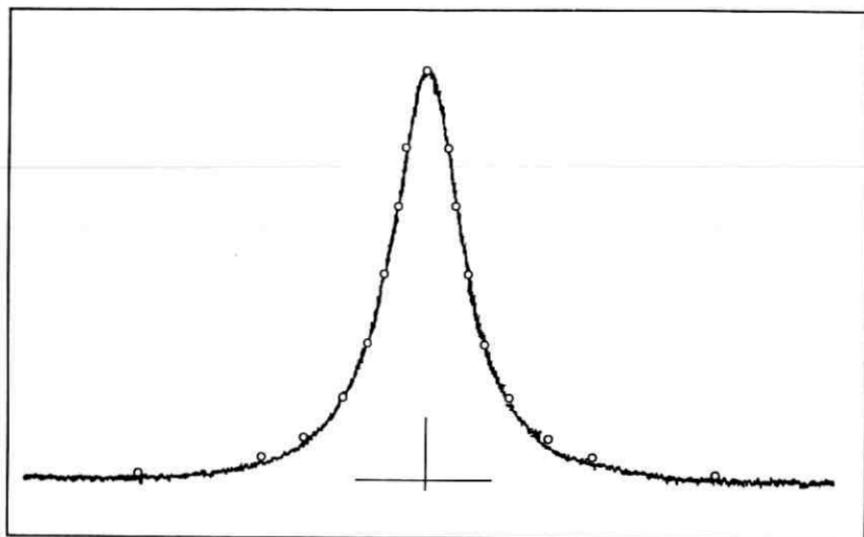


Fig. 15 — Far-field diffraction pattern of the beam scattered from a 245 mc/sec acoustic wave in water. The open circles correspond to a true Lorentzian matched to the peak and half-power points of the experimental curve.

The spot size of the Gaussian light beam was several times wider than the acoustic decay distance.

Near-field traces of scattered Gaussian, half-Gaussian, and uniform beams were obtained and gave results in qualitative agreement with the results of the previous section. No quantitative measurements were made.

Quantitative measurements were performed on the far-field pattern. A typical result is illustrated in Fig. 15 and was obtained with a half-Gaussian beam with a spot size in excess of one centimeter. The open circles are for a true Lorentzian matched to the experimental curve at the peak and half-power points. Note that the wings of the experimental curve are slightly less than the true Lorentzian value, as might be expected from the edge effect. The measured angular width [using (36)] agrees with the measured value of α within 5 per cent.

2.4 Conclusion

A technique for probing acoustic beams in optically transparent materials has been described. The technique is based on a Fourier transform relationship between the intensity distribution in the cross section of the acoustic beam and the angular dependence of the optical-acoustic interaction. It allows unequivocal determination of the volume

acoustic loss or phonon lifetime independent of the transducer and other boundary effects, precise determination of the average direction of the acoustic beam, and observation of the far-field diffraction pattern of the acoustic beam. It also determines the angular response of the transducer. When there is reason to believe that there is no phase shift in the cross-section, or if there is knowledge concerning the phase shift, the spatial distribution can be determined from the inverse Fourier transform of the measured angular distribution. Experiments illustrating and verifying these relationships have been described. Some of the far-field results for the scattered light intensity have relevance to optical beam deflection devices.⁷

ACKNOWLEDGMENT

The authors are greatly indebted to N. F. Foster, who fabricated the transducers used in these experiments.

APPENDIX

The amplitude of the incident Gaussian beam is given by

$$E_0(x,y) = \exp -\frac{1}{2}r^2/w^2.$$

Fig. 12 shows that for a beam incident at angle Θ with respect to the y axis

$$r = y \sin \Theta + x \cos \Theta$$

so that

$$E_0(x,y) = \exp -\frac{1}{2} \frac{(y \sin \Theta + x \cos \Theta)^2}{w^2}.$$

The scattered wave originating from $(x_0, 0)$ travels along the line $x = x_0 + y \tan \Theta$. The acoustic wave amplitude is given by $\exp -\alpha x$. The total scattered amplitude associated with the point $(x_0, 0)$ on the incident light beam can be written

$$E_1(\Theta, x_0) \propto \int_0^L dy \exp -\alpha(x_0 + y \tan \Theta) - \frac{1}{2} \frac{[y \sin \Theta + (x_0 + y \tan \Theta) \cos \Theta]^2}{w^2} \quad (42)$$

following the discussion in Section II. Multiplying out and completing the square in the exponential yields

$$E_1(\Theta, x_0) \propto \exp -\alpha x_0 - \frac{1}{2} \frac{x_0^2}{w^2} \cos^2 \Theta + \frac{w^2}{8 \sin^2 \Theta} \left(\alpha \tan \Theta + \frac{x \sin 2 \Theta}{w^2} \right)^2 \times \int_0^L dy \exp - \frac{2 \sin^2 \Theta}{w^2} \cdot \left[y + \left(\alpha \tan \Theta + \frac{x_0 \sin 2 \Theta}{w^2} \right) \frac{w_0^2}{4 \sin^2 \Theta} \right]^2. \quad (43)$$

With the substitution

$$y' = \left(y + \frac{1}{4} \left[\frac{\alpha w}{\cos \Theta} + \frac{2 x_0}{w} \cos \Theta \right] \frac{w}{\sin \Theta} \right) \frac{\sqrt{2} \sin \Theta}{w}$$

(43) becomes

$$E_1(\Theta, x_0) \propto \exp -\alpha x_0 - \frac{1}{2} \frac{x_0^2}{w^2} \cos^2 \Theta + \frac{1}{8} \left(\frac{\alpha w}{\cos \Theta} + \frac{2 x_0 \cos \Theta}{w} \right)^2 \times \frac{w}{\sqrt{2} \sin \Theta} \int \frac{\sqrt{2}}{4} \left[\frac{\alpha w}{\cos \Theta} + \frac{2 x_0 \cos \Theta}{w} \right] \frac{w}{\sin \Theta} \frac{\sqrt{2} \sin \Theta}{w} \exp -y'^2 dy'. \quad (44)$$

This reduces to (30) upon substitution of

$$X = \sqrt{2} \left(\frac{L}{w} \right) \sin \Theta$$

$$Y = \frac{1}{4} \sqrt{2} \left(\frac{\alpha w}{\cos \Theta} + 2 \left(\frac{x_0}{w} \right) \cos \Theta \right).$$

REFERENCES

1. Born, M., and Wolf, E., *Principles of Optics*, Pergamon Press, New York, 1959. Ch. 12 contains extensive bibliography and description of the earlier work.
2. Mason, W. P., *Piezoelectric Crystals and Their Application to Ultrasonics*, D. Van Nostrand Co., Princeton, N.J., 1950.
3. Fitch, A. H., and Dean, R. E., A Focusing Ultrasonic Delay Line, 1964 Symposium on Sonics and Ultrasonics, Santa Monica, Calif.
4. Quate, C. F., Shaw, H. J., Tien, P. K., Wilkinson, C. D. W., and Winslow, D. K., Diffraction of Laser Light with Hypersound, M.L. Report No. 1172, Microwave Laboratory, Stanford University, June, 1964.
5. Rigden, J. D., and Gordon, E. I., The Granularity of Scattered Optical Maser Light, Proc. IRE, 50, Nov., 1962, p. 2367.
6. Herzfeld, K. F., and Litovitz, T. A., *Absorption and Dispersion of Ultrasonic Waves*, Academic Press, New York, 1959.
7. Cohen, M. G., and Gordon, E. I., Electro-Optic [KTa_zNb_{1-z}O₃(KTN)] Gratings for Light Beam Modulation and Deflection, Appl. Phys. Lett., 5, 1964, p. 181.

Contributors to This Issue

VÁCLAV E. BENEŠ, A.B., 1950, Harvard College; M.A., and Ph.D., 1953, Princeton University; Bell Telephone Laboratories, 1953-. Mr. Beneš has been engaged in mathematical research on stochastic processes, traffic theory, and servomechanisms. In 1959-60 he was visiting lecturer in mathematics at Dartmouth College. He is the author of *General Stochastic Processes in the Theory of Queues* (Addison-Wesley, 1963). Member, American Mathematical Society, Association for Symbolic Logic, Institute of Mathematical Statistics, SIAM, Mind Association and Phi Beta Kappa.

CHRISTOPH B. BURCKHARDT, Dipl.-Ing., 1959, Dr. sc. techn., 1963, Swiss Federal Institute of Technology; Bell Telephone Laboratories, 1963-. He has been engaged in the analysis of varactor frequency multipliers and the large-signal behavior of parametric amplifiers. He is currently working on optical storage of information. Member, IEEE.

MARTIN G. COHEN, A.B., 1957, Columbia College; M.A., 1958, Ph.D., 1963, Harvard University; Bell Telephone Laboratories, 1964-. He has been engaged in research in optical modulation techniques since joining Bell Laboratories. Member, American Physical Society, Phi Beta Kappa and Sigma Xi.

EUGENE I. GORDON, B.S., 1952, City College of New York; Ph.D., 1957, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1957-. He has been engaged in research in plasma physics, microwave tubes, lasers, and light modulation. He is presently Head of the Optical Device Department. Member, American Physical Society, Phi Beta Kappa, Sigma Xi and IEEE.

ROBERT W. LUCKY, B.S.E.E., 1957, M.S.E.E., 1959, Ph.D., 1961, Purdue University; Bell Telephone Laboratories, 1961-. Mr. Lucky has been concerned with various analysis problems in the area of digital data communications. Member, IEEE, Sigma Xi, Tau Beta Pi and Eta Kappa Nu.

LEWIS C. THOMAS, B.E.E., Cornell University, 1949; M.S. in E.E., c.s.l., 1958, Newark College of Engineering; Bell Telephone Laboratories, 1949-. He has worked on the Nike missile systems, pulse code modulation systems, and data transmission systems. His recent work has included attitude and orbital mechanics studies for Project Telstar and communication satellite system studies. Member, IEEE, AIAA, Royal Astronomical Society of Canada, Eta Kappa Nu and Epsilon Pi Tau.

B.S.T.J. BRIEFS

On the Simultaneous Measurement of a Pair of Conjugate Observables

By E. ARTHURS and J. L. KELLY, JR.

(Manuscript received December 16, 1964)

A precise theory of the simultaneous measurement of a pair of conjugate observables is necessary for obtaining the classical limit from the quantum theory, for determining the limitations of coherent quantum mechanical amplifiers, etc. The uncertainty principle, of course, does not directly address this problem, since it is a statement about the variances of two hypothetical ideal measurements. We will adopt the approach that there exist instantaneous inexplicable ideal measurements of a single observable. Just as von Neumann¹ uses an ideal measurement together with an interaction to explain an indirect observation, we use ideal measurements together with interactions to explain the simultaneous measurement of an observable and its conjugate.

The joint measurement described below is complete in the sense that all pertinent past history is subsumed under the meter readings. A precise formula for the joint probability distribution of the results of the measurements is given. The variances given by these distributions satisfy an inequality like the uncertainty principle but with an extra factor or two. It is also shown that this inequality governs any conceivable joint measurement. The single measurement of an observable is a limiting case of the joint measurement when the variance of one of the measured variables is allowed to approach infinity.

Since we are trying to measure two observables, we will introduce two meters, that is, two one-dimensional systems which will be coupled to the object system. Since the two meter positions commute, we can make ideal simultaneous measurements of them. Our interpretation will be that these two measurements will constitute a simultaneous measurement of the two noncommuting observables of the object system. We will see that we cannot let the strength of our interaction become infinite, unlike the indirect measurement of Ref. 1, but must adjust it to a certain critical value. We will find that after the interaction and the measurement of the meter values that the system is left in a state which is completely determined by the meter readings and a certain parameter

which we will call the "balance," having to do with how near our joint measurement is to either of two ideal measurements.

To effect the desired measurement, the coupling must be described by the following Hamiltonian:

$$H_{\text{int}} = K(qP_x + pP_y) \quad (1)$$

where p, q are the positions and momentum we wish to be measured and P_x, P_y are the momenta of the two single-degree-of-freedom systems we are using for meters. The positions of the two meters will be x and y . In addition, we require that the meters initially be in the states $M(x)$ and $N(y)$, where

$$M(x) = \left(\frac{2}{\pi b}\right)^{\frac{1}{2}} e^{-x^2/b} \quad (2)$$

$$N(y) = \left(\frac{2b}{\pi}\right)^{\frac{1}{2}} e^{-by^2}$$

and b is the "balance." The strength of the interaction K is assumed to be sufficiently large that other terms in the Hamiltonian can be ignored. Hence the Schrödinger equation will be ($\hbar = 1$)

$$\frac{\partial \varphi}{\partial t} = -K \left(q \frac{\partial \varphi}{\partial x} - i \frac{\partial^2 \varphi}{\partial q \partial y} \right) \quad (3)$$

with the initial condition

$$\varphi(q, x, y, 0) = F(q)M(x)N(y) \quad (4)$$

where $F(q)$ is the state of the system to be measured and the system and the two meters are assumed to be independent prior to the interaction. Equation (3) is solved by Fourier transforming on y . The solution is

$$\varphi(q, x, y, t) = \int_{-\infty}^{\infty} F(q - wtk) M(x - qtk + \frac{1}{2}wt^2k^2) \cdot \frac{\exp - (w^2/4b)}{(4\pi b)^{\frac{1}{2}}} \exp(iwy) dw. \quad (5)$$

To obtain the results we desire it is necessary to make ideal measurements of x and y at $t = 1/K$. In the following, t will be assumed to be equal to $1/K$ and the time will be suppressed in the notation.

The joint probability distribution for the commuting observables x and y , $P(x, y)$ is given, of course, by

$$P(x, y) = \int_{-\infty}^{\infty} |\varphi|^2 dq. \quad (6)$$

Then, using (5), we have

$$P(x,y) = \left(\frac{1}{4\pi^3 b}\right)^{\frac{1}{2}} \left| \int_{-\infty}^{\infty} F(q) \exp - \left[\frac{1}{2b} (x - q)^2 \right] \cdot \exp - (iqy) dq \right|^2 \quad (7)$$

or, if $G(p)$ is the momentum wave function of the system

$$P(x,y) = \left(\frac{b}{4\pi^3}\right)^{\frac{1}{2}} \left| \int_{-\infty}^{\infty} G(p) \exp - \left[\frac{b}{2} (y - p)^2 \right] \cdot \exp - (ixp) dp \right|^2 \quad (8)$$

That is, the joint probability distribution is the Fourier spectrum of the wave function multiplied by a Gaussian window whose width is related to the balance of the measurement.

The new wave function for the system after the measurement is given by substituting the results of the meter readings in the wave function and renormalizing. If we denote the measured value of x by x_m and the measured value of y , y_m , the new state of the system is given by

$$F'(q) = \frac{\varphi(q, x_m, y_m)}{\int |\varphi(q, x_m, y_m)|^2 dq} = \left(\frac{1}{\pi b}\right)^{\frac{1}{2}} \cdot \exp - \left[\frac{1}{2b} (q - x_m)^2 + iqy_m \right]. \quad (9)$$

Notice that the measurement is complete, in that the state of the system after the measurement is dependent only on the meter readings and not otherwise on the state of the system before measurement. We also notice that the system is left in a minimum Gaussian packet after the measurement, with mean position x_m and mean momentum y_m , which is an intuitively satisfying result.

From (7) or (8) it is easy to verify that the expected value of x is equal to the expected value of q before the interaction and that the expected value of y is equal to the expected value of p before the interaction. The variances of x and y are related to the variances of q and p before the interaction by

$$\begin{aligned} \sigma_x^2 &= \sigma_q^2 + b/2 \\ \sigma_y^2 &= \sigma_p^2 + 1/2b. \end{aligned} \quad (10)$$

Hence the variances are individually larger than those of the wave function $F(q)$ due to the disturbances caused by the joint measurement. From

(7) and (10) it can be seen that in the limit $b \rightarrow 0$ distribution of x is the same as that of an ideal measurement of position and the system is left in an eigenstate of position. Similarly, if $b \rightarrow \infty$ we have an ideal momentum measurement and the system is left in an eigenstate of momentum.

From (10) and the uncertainty principle

$$\sigma_q \sigma_p \geq 1/2 \quad (11)$$

we can deduce

$$\sigma_x \sigma_y \geq 1 \quad (12)$$

which is the proper uncertainty principle for the joint measurement. The minimum can actually be met when $F(q)$ is a minimum Gaussian packet and the balance b is suitably adjusted. It is interesting that, when $F(q)$ is a minimum Gaussian packet, (8) shows that the meter readings are distributed as independent Gaussian random variables.

We will now show that the bound expressed by (12) is valid for any joint measurement that meets certain reasonable requirements.

Let us consider a joint measurement from a more general point of view. As before, we will have a meter (a system with at least two degrees of freedom) interact with our system. The initial wave function for the meter plus system will be the product of the system wave function $F(q)$ and the meter function $M(w_1 w_2, \dots)$. After allowing the interaction to proceed for t seconds we will measure two observables, say $x(t)$ and $y(t)$, which will hopefully measure the system position and momentum. In the Heisenberg representation we may write without loss of generality

$$\begin{aligned} x(t) &= q(0) + A \\ y(t) &= p(0) + B. \end{aligned} \quad (13)$$

If we normalize with a scale factor of unity on both measurements, it is natural to require that the expectations of $x(t)$ and $y(t)$ satisfy

$$\begin{aligned} \langle x(t) \rangle &= \langle q(0) \rangle \\ \langle y(t) \rangle &= \langle p(0) \rangle \end{aligned} \quad (14)$$

uniformly for *all* initial states of the system, i.e., for all $F(q)$. This implies that

$$\begin{aligned} \langle A \rangle &= 0 \\ \langle B \rangle &= 0 \end{aligned} \quad (15)$$

identically for all $F(q)$. From this and the fact that the initial wave function for the system plus meter factors, it can be shown

$$\begin{aligned}\langle qA \rangle &= \langle Aq \rangle = \langle Bq \rangle = \langle qB \rangle = 0 \\ \langle pA \rangle &= \langle Ap \rangle = \langle Bp \rangle = \langle pB \rangle = 0.\end{aligned}\tag{16}$$

Secondly, we require that $x(t)$ and $y(t)$ commute so that they may be simultaneously measured. From this and (13) we have

$$[B, A] = [q, p] + [q, B] + [A, p].\tag{17}$$

Squaring both sides of (17) and taking expectations, it follows from (16) that

$$\langle -[A, B]^2 \rangle \geq 1\tag{18}$$

which implies

$$\langle A^2 \rangle \langle B^2 \rangle \geq 1/4.\tag{19}$$

Using (13) and (16), we obtain

$$\sigma_x^2 \sigma_y^2 = \sigma_p^2 \sigma_q^2 + \langle A^2 \rangle \langle B^2 \rangle + \sigma_q^2 \langle B^2 \rangle + \sigma_p^2 \langle A^2 \rangle\tag{20}$$

where $\sigma_x^2 = \langle (x - \langle x \rangle)^2 \rangle$, etc.

From (11) and (19) it follows that

$$\sigma_x \sigma_y \geq 1\tag{21}$$

which is the desired result.

REFERENCES

1. von Neumann, J., *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton, N. J., 1955.

