The Use of Polarization Fading of Satellite Signals To Study the Electron Content and Irregularities in the Ionosphere[1, 2]

C. Gordon Little and Robert S. Lawrence

(January 12, 1960)

A procedure is described for using the Faraday-rotation fading of a satellite radio signal to measure the ionospheric electron content per unit column up to the height of the satellite. At frequencies as low as 20 Mc the rotation of the plane of polarization cannot be assumed to be proportional to $\int NB \cos \theta \, dl$ along the line of sight. The simplifying assumptions implied by this expression are avoided, and full account is taken of ionospheric refraction, using the collision-free form of the Appleton-Hartree equation. Results based on observations of 1958 Delta 2 are presented. The subsatellite electron contents have been derived throughout the satellite passes for heights both above and below the F-peak; the latter compare well with values derived from simultaneous ionograms. The method also permits the study of large-scale irregularities in electron content. Such irregularities, having lateral dimensions of a few hundred kilometers and fractional deviations in subsatellite electron content of about 0.01, have been detected. Our observations suggest that satellite polarization studies offer important advantages over other methods of investigating these irregularities.

1. Introduction

The study of the ionosphere by means of the Faraday rotation of the plane of polarization of a radio wave has developed from the early moon-radar work of Murray and Hargreaves [1]. These and other authors have shown that, after making certain assumptions, one may write

$$\Phi = \frac{A}{f^2} \int_l NB \cos \theta \, dl \qquad (1)$$

where

$\Phi =$ the rotation of the plane of polarization,
$A =$ a constant,
$f =$ frequency of radio wave,
$N =$ number density of electrons,
$B =$ magnetic field strength,
$\quad \theta =$ angle between magnetic field and direction of propagation,
$dl =$ element of path along the line of sight l.

It has been customary to rewrite eq (1) in the form

$$\Phi = \frac{A}{f^2} [B \cos \theta \sec i]_{\text{avg}} \int_0^s N \, dh \qquad (2)$$

where i is the zenith angle of the line of sight, and a mean or "effective" value of $B \cos \theta \sec i$ is taken along the ionospheric part of the path.

Equation (2) shows that, within the limits of accuracy imposed by the approximations used in its derivation, an experimental measurement of Φ permits calculation of the electron content in a vertical column of the ionosphere. In practice, Φ is usually indeterminate by $n\pi$, and additional information is normally required before the value of $\int_0^s N \, dh$ can be derived unambiguously.

The approximations implicit in eq (2) are as follows:

(1) The observing frequency, f, is very much larger than the maximum plasma frequency, f_c, along the line of sight.

(2) The observing frequency is very much greater than the gyromagnetic frequency, f_H.

(3) The quasi-longitudinal approximation of the Appleton-Hartree equation may be used.

(4) The zenith distance, i, is small and the ionospheric thickness is very small compared with the radius of the earth. (These conditions restrict the variation of B, $\cos \theta$ and $\sec i$ along the line of sight and permit the use of an average value of each quantity along the ionospheric part of the path. They also minimize the errors which arise due to the neglect of ionospheric refraction. In the presence of such refraction, the radio energy does not traverse the single straight line assumed in eq (2); instead the ordinary and extraordinary waves traverse two different refracted paths.)

(5) The ionospheric electron density contours are spherically stratified.

In view of the above limitations, the moon-radar studies of ionospheric electron content have usually been made at frequencies in excess of 100 Mc and at fairly high angles of elevation.

The advent of satellite-borne radio transmitters has offered a new opportunity for Faraday-rotation studies of the ionosphere [2]. It is important to note that satellite studies, in principle at least, have the advantage that the electron content of the ionosphere can be measured over a wide range of latitudes from a single observing site, within the course of the few minutes duration of a satellite passage. In order to make full use of this advantage, however, it is necessary to remove the restriction to low zenith angles; it is also desirable to use frequencies relatively close to the ionospheric critical frequency in order to improve the sensitivity of the observations.

The failure of eq (2) at low frequencies and at large zenith angles is strikingly demonstrated by simultaneous observations of the 20- and 40-Mc signals from the Soviet satellites. Equation (2) predicts that the ratio of the Faraday rotation rates on the two frequencies shall be 4.0; in fact ratios of up to 5.0 have been observed, and on occasion, for a short period of time the sense of rotation has even been opposite on the two frequencies.

The purpose of this paper is to present a method of analysis whereby the assumptions implicit in eq (2) are eliminated. The method permits the deduction of accurate values of $\int_0^s N\,dh$ throughout a satellite pass, at frequencies as low as 20 Mc.

The main part of the paper opens with a description of the techniques used in recording the Faraday-rotation data. This is followed by an outline of the method of analysis; results of the application of the method to a low and a high pass of 1958 Delta 2 (Sputnik III) are given. The accuracy and limitations of the method are then examined. The paper concludes with a discussion of the large-scale ionospheric irregularities revealed by the analysis, and of the sensitivity and value of the Faraday-rotation technique for such studies.

2. Observational Methods

The satellite-observing facility at the Table Mesa field site (lat. 40°7.7′ N, long. 105°14.3′ W) of the Central Radio Propagation Laboratory was constructed in the spring of 1958. Interferometer, Doppler, and polarization measurements have been made each day since May 15, 1958, on the close passes of 1958 Delta 2, using a variety of antennas, baseline lengths, and baseline orientations.

The Faraday rotation records which form the observational data for this paper were taken using horizontal, linearly polarized 20-Mc dipoles mounted λ/3 above the ground. The antennas were connected via coaxial transmission lines to standard commercial communication receivers which were operated with the receiver AVC on. The AVC voltages were recorded on multichannel Sanborn pen recorders at chart speeds of 1 cm/sec. In the absence of AVC the receiver output would have varied quasi-sinusoidally as the plane of polarization rotated relative to the antenna. AVC was used in order to obtain an essentially logarithmic input-output receiver characteristic; the normal quasi-sinusoidal output waveform was thereby replaced by a waveform having broad maximums and very narrow, well-defined nulls. These nulls were often sufficiently well defined to permit their timing to an accuracy of 0.1 sec.

An alternative receiver output presentation has also been used. Two receivers having equal gain were connected to two similar antennas, one oriented N–S, the other E–W. The AVC outputs of the two receivers were subtracted and the resultant voltage recorded on the pen recorder. Since the receiver outputs for the two orthogonal antennas may be (approximately) expressed as log sin ωt and log cos ωt respectively (where ω=angular rate of rotation of the plane of polarization), their difference is equal to log (sin ωt/cos ωt)=log (tan ωt). The resultant waveform is nearly triangular, with equal positive and negative slopes and cusps at the null-times on each antenna. This form of presentation was found advantageous when the satellite signal was fading irregularly, or when the fading rate was so fast as to merge with the beeping of the satellite signal.

A third method, used occasionally, was to record the relative phase of the radio signals received on circularly polarized antennas of opposite sense of rotation. This method has the advantage of providing continuous information on the orientation of the plane of polarization, rather than information on null-times only. The technique was not found useful on 1958 Delta 2, primarily owing to the intermittent, beeping character of the signal.

3. Method of Analysis

Our method for the quantitative interpretation of the observed Faraday-rotation data is described below. The essential feature is the computation of the expected time variation of the specific rotatory power, $C(t)$, of the electrons along the line of sight, as the satellite moves along a known path through a model ionosphere. $C(t)$ is defined by

$$C(t) = \frac{\Phi_M(t)}{I_M(t)} \qquad (3)$$

and is a measure of the efficiency of the electrons below the satellite in producing Faraday rotation. $\Phi_M(t)$ is the variation, with time, of the total Faraday rotation of the satellite signal as it traverses the model ionosphere en route to the observer, and $I_M(t)$ is the variation with time of the subsatellite electron content, $\int_0^s N\,dh$, in the model. $\Phi_M(t)$ is computed by a ray-tracing program, and $I_M(t)$ is readily obtained by an integration of the model ionosphere, hence $C(t)$ can be computed.

In order to derive $C(t)$ correctly, a moderately accurate ionospheric model is desirable, although as shown in section 5, high accuracy is not required. This fortunate feature arises from the fact that reasonable changes in the model affect $\Phi_M(t)$ and

$I_M(t)$ in almost the same proportion, and their ratio is therefore essentially unchanged.

Having determined $C(t)$ for the model in this manner, we make use of the observed values of Faraday rotation, $\Phi(t)$ to determine the true ionospheric electron content by writing

$$I(t) = \frac{\Phi(t)}{C(t)} \tag{4}$$

The observed values of $\Phi(t)$ are, in general, uncertain by an additive constant, $n\pi$; the manner by which this uncertainty may be reduced is discussed in section 4.

The method is now examined in greater detail; the reader will find it helpful to refer to figure 1 as each new step is discussed.

Orbital elements of the satellite were used to compute its position, relative to the observing station, at each of several selected times during the passage. The elements used were based on those of the Smithsonian Astrophysical Observatory or upon predictions made by the Vanguard Computing Center. The latter, being predictions, were much less accurate than the former but, if the time of closest approach is adjusted to agree with local observations, the remaining errors have been found to have slight effect on the derived values of I.

For heights up to F-peak, spherically stratified ionospheric models were used, based upon true-height analysis of local ionosonde data. The profiles above F-peak were obtained by assuming them to be those of a Chapman layer. This approach is a convenient one, as the Chapman function has the proper slope at the F-peak and asymptotically approaches a decaying exponential function at great heights. After fitting the function at F-peak, only one adjustable parameter remains, the asymptotic exponential decay constant. We shall refer to this constant as the "electron scale height." It should be noted that the Chapman function is here used only as a convenient mathematical model, without any regard to its physical basis. If the values of I obtained from a satellite which is above the F-peak differ significantly from the corresponding values of I_M, the electron scale height may be adjusted until agreement is reached.

The assumption of spherical stratification can be removed by the inclusion of a horizontal gradient in the ionosphere model. One method for the determination of such a gradient from spaced ionograms is illustrated in figure 2. Here values of f_0F2 have been plotted from simultaneous observations over a wide area, and contours of f_0F2 drawn. The values of critical frequency thus deduced along the satellite track (and assumed to be at the same true height) provide factors by which the electron-density profile at Boulder must be multiplied to approximate the actual ionosphere along the track. The necessity for including horizontal gradients adds no particular difficulty to the ray-tracing procedure, but it provides another adjustable parameter which must be determined, and thus it weakens the determination of electron scale height above the F-peak.
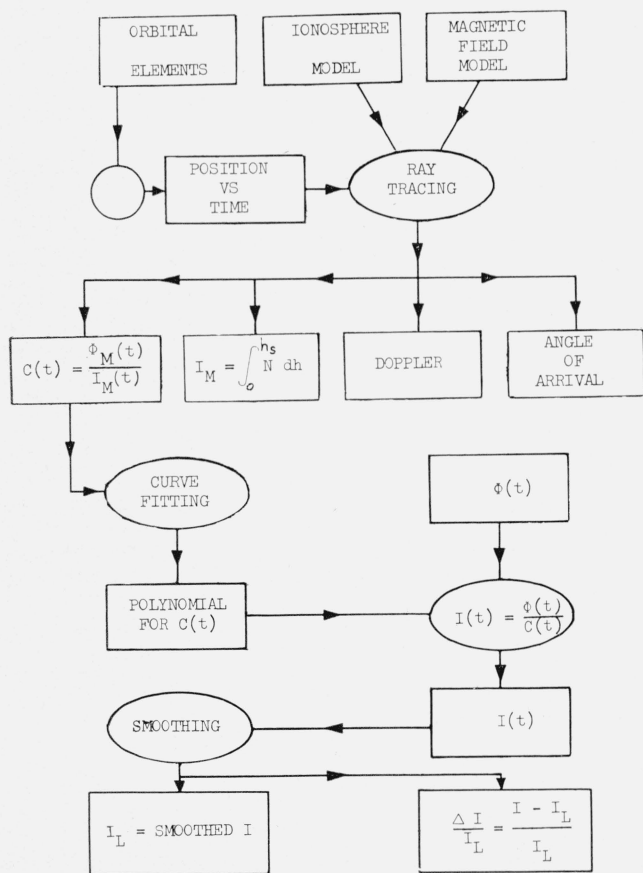

FIGURE 1. *Method of analysis of satellite Faraday rotation records.*

The circles or ovals represent processes carried out with a digital computer; the rectangular boxes represent input or output data for these processes.
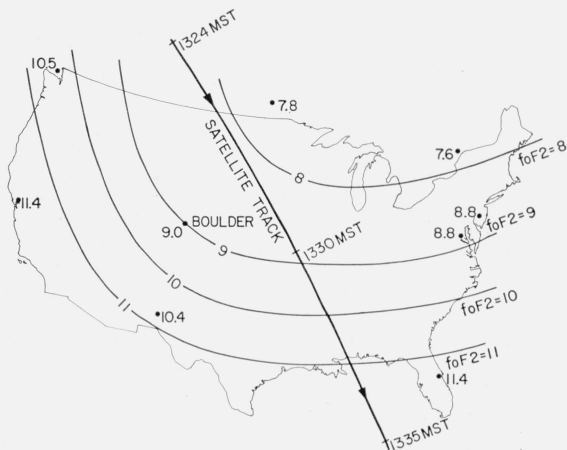

FIGURE 2. *Geographical variation of f_0F2, used to estimate the horizontal ionospheric gradient, September 17, 1958, 1330 m.s.t.*

The geomagnetic field is commonly represented by the field of a dipole of appropriate strength and orientation at the center of the earth. When, as in the present application, the field is needed over only a limited portion of the earth, a significant improvement can be made by adjusting the strength and orientation of the dipole to best fit the observed field in that region. Such a "modified dipole" field has been used for some of the results presented here. A more accurate, but much more laborious method is to add higher-order terms to the spherical-harmonic representation of the field. This method has been used with terms up to order six, but only for comparison with the modified dipole results because it involves an impractically long calculation. When used in ray tracing, the 48-term polynomial for the magnetic potential must be solved four times to calculate the vector magnetic field. For each position of the satellite, this entire operation is repeated at dozens of points along each ray, and must be done not only for the final rays but for all the trial rays which fail to connect the satellite and the observer. The relative accuracy of results based on the modified dipole field and the complete spherical-harmonic representation is discussed in a later section.

Once the satellite motion is known and models have been chosen for the ionospheric electron-density profile and for the geomagnetic field, the specific rotatory power, $C(t)$, of the ionosphere can be obtained by ray tracing. Since C depends upon the direction of the geomagnetic field relative to the ray and on the zenith angle of the ray, it obviously varies with time during the passage of the satellite. In obtaining Φ_M, and hence $C(t)$, we have used a variety of ray-tracing programs.

The simplest program we have used evaluates $\int_o^s N B \cos \theta \sec i \, dh$ along the straight line from the observer to the satellite. This is an improvement over the use of eq (2) in that it eliminates the assumption of a thin ionosphere and, to some extent, permits the use of larger zenith angles and horizontal ionospheric gradients. The high-frequency and quasi-longitudinal approximations remain, however, and the effects of refraction are ignored.

A second program, designed to overcome the difficulties of the high-frequency approximation, evaluates the integral $2\pi/\lambda \int_0^s (\mu_0 - \mu_e) \, dl$ along the straight line from observer to satellite. Here μ_o and μ_e are the local refractive indices of the ionosphere for ordinary and extraordinary waves traveling along the line of sight. Half of this integrated difference between the ordinary and extraordinary phase path lengths gives Φ_M. I_M is determined by simple integration of $N \, dh$ along the straight line. Since this second program uses the complete collision-free form of the Appleton-Hartree equation (see for example, ref [3]) to evaluate μ_0 and μ_e, the high-frequency approximation is removed. For nearly transverse propagation the Faraday rotation is critically dependent upon θ, so it is meaningless to

speak of removing the quasi-longitudinal approximation as long as refraction is neglected.

A third program, involving considerably more computer time, includes most of the effects of refraction. This program evaluates $2\pi/\lambda \int_0^s (\mu_0 - \mu_e) \, dl$ along the refracted path which the radio ray would follow in the absence of the geomagnetic field. This program assumes that the ordinary and extraordinary rays traverse the same refracted path. It works well in the quasi-longitudinal part of the sky but is quite unsatisfactory near the quasi-transverse region, where the Faraday rotation is critically θ-dependent.

A final program has been developed for use in this critical portion of the sky. It is a complete ray-tracing program which evaluates separately the phase paths, P_o and P_e, of the ordinary and extraordinary rays, along their independent refracted paths. This full ray-tracing is accomplished by using Snell's law at the interfaces of a series of elementary spherical shells, each considered to possess constant refractive index. Referring to figure 3, let O_i be the intersection of the ray with the interface between spherical shells i and $i+1$. $\mathbf{N_i}$ is the unit vector normal to the interface and $\mathbf{W_i}$ is the unit vector in the direction of the wave normal in shell i. $\mathbf{B_i}$ is the geomagnetic field. The electron densities, N_i and N_{i+1}, are known in each shell. The refractive index, μ_i, is known, but μ_{i+1} is not known since it depends upon θ_{i+1}. A first approximation to μ_{i+1} may be obtained by setting $\theta_{i+1} = \theta_i$. Snell's law then gives a second approximation to θ_{i+1}, and hence to μ_{i+1}. Two or three iterations are required to determine μ_{i+1} and θ_{i+1} with sufficient accuracy, and thus to determine the direction $\mathbf{W_{i+1}}$ of the new wave normal. As discussed by Bremmer [4], the new ray, $\mathbf{R_{i+1}}$, does not coincide with $\mathbf{W_{i+1}}$ but is separated from it by an angle α. $\mathbf{R_{i+1}}$ lies in the plane of $\mathbf{N}$ and $\mathbf{B}$ and so is not necessarily in the plane of incidence. An extension of $\mathbf{R_{i+1}}$ through the shell defines O_{i+1} at the second interface and completes the ray tracing through shell $i+1$.
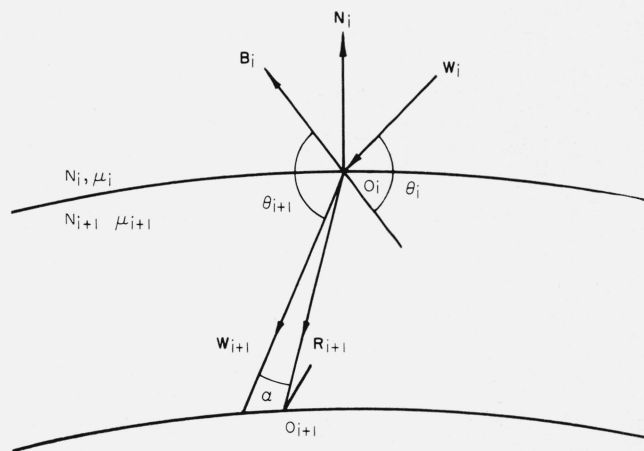

FIGURE 3. *Detail of ray tracing through one spherical shell.*

When a ray has been traced from the satellite to the ground it will not, in general, arrive at the observer. It is then necessary to make a correction to the direction of departure from the satellite, and trace another ray. Four to eight such attempts, depending upon the zenith angle, are usually sufficient to discover the ray which, with sufficient accuracy, connects the satellite to the observer. The accuracy required is such that the small difference in phase paths, $P_o - P_e$, is known to an accuracy of one percent. (This typically requires the determination of each phase path length to about one part in 10^6.) Shell thicknesses as small as 10 km are generally required. In extreme cases, the IBM 650 computer may require from 30 to 60 min to complete the process for a single position of the satellite. Accordingly, this full ray-tracing program is used only near the quasi-transverse regions of the sky, or to evaluate the accuracy of the faster, more approximate programs.

The ray-tracing procedure just described is strictly accurate only when the surfaces of constant refractive index coincide with the spherical shells. In fact, it has proved to be sufficiently accurate in the case of all reasonable horizontal gradients in the ionosphere. In these cases, the electron density of each shell must be given in terms of the geographical location of the ray, but comparison with a more rigorous method has shown that the interfaces between shells may still be considered to be normal to the radius vector from the center of the earth. A detailed description of the ray-tracing program is in preparation.

Byproducts of the ray-tracing procedure are the Doppler frequency shift and the angle of arrival. The observed Doppler shift depends not only upon the radial velocity of the satellite, but also upon the local electron density and upon the rate of change of integrated electron density along the ray path. These ionospheric effects are automatically included in the results of the ray tracing. Similarly, the ray-tracing program evaluates the difference in angle of arrival between the ordinary and extraordinary rays, the total bending of each, and the spatial separation between each ray path and the line of sight.

It is clearly impractical to evaluate $C(t)$ at the time of each observed Faraday-rotation null. Since its variation depends only upon the geometrical position of the satellite and upon the geomagnetic field along the ray, $C(t)$ is a smoothly varying function of time and can be represented very well by a simple polynomial. Typically, $C(t)$ has been evaluated by ray tracing at nine equally spaced times throughout the satellite passage, and the method described by Anderson [5] has been used to fit, by least squares, a fifth-degree polynomial to these nine points. In every case the polynomial has agreed with the nine given values to within a small fraction of one percent.

Once a polynomial expression has been obtained for $C(t)$, any given value of $\Phi(t)$ is readily converted to a corresponding value of I by means of eq (4). Successive Faraday-rotation nulls which appear on the recordings of the satellite signal correspond to values of Φ which differ by π radians. It is necessary only to observe the times at which such nulls occur, and then to assign to some one null an assumed value of Φ, in order to obtain a list of values of $\Phi(t)$. The initial value of Φ is usually selected to be equal to the corresponding Φ_M from the ray-tracing program. A method of improving this initial estimate is demonstrated in the next section.

In some records the direction of rotation of the plane of polarization is found to reverse, i.e., $d\Phi/dt$ changes sign. It is important to recognize such events on the record, and this can be done most easily by noting the abrupt change in relative phase of the fading on two orthogonal linearly-polarized antennas.

Details of the relatively small fluctuations which occur in $I(t)$ become evident only by separating them from the gross value of I. $I(t)$ was therefore smoothed by a running average of suitable duration, to give $I_L(t)$. The duration of this smoothing interval was chosen to equal the time required for the portion of the ray which is 300 km above the ground to move through a distance of 400 km. The exact duration of this interval is therefore a function of the height of the satellite and of its velocity. The fractional irregularities in $I(t)$ are given by

$$\frac{\Delta I}{I_L} = \frac{I - I_L}{I_L}$$

This process of taking the difference from a smoothed value is in fact a spatial filter. It has the property of rejecting all Fourier components with wavelengths much above 400 km and has relatively little effect upon the amplitude of components smaller than the smoothing interval.

4. Some Preliminary Results of the Method

Figure 4 is a plot of the derived subsatellite electron content of the ionosphere during the northbound pass number 1073 of 1958 Delta 2. This
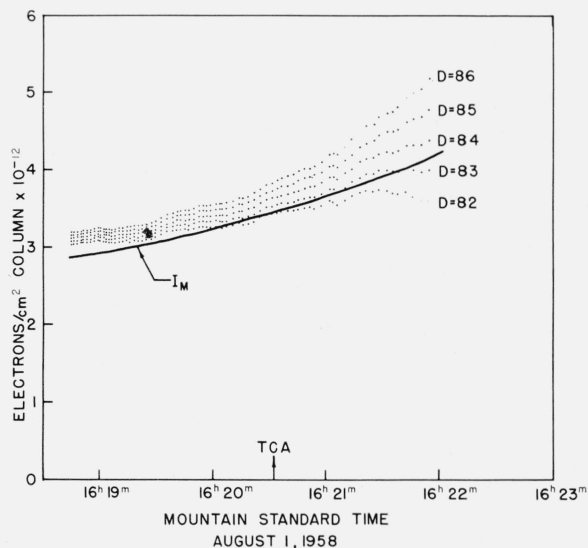

FIGURE 4. *Comparison of* I_M *with Faraday rotation data for five values of* D.

pass took place on the afternoon of August 1, 1958; perigee occurred with the satellite near the southern horizon, so the satellite was rising slowly as it passed almost overhead at a height of about 232 km.

The model of the ionosphere used in the analysis of figure 4 was obtained by a linear interpolation between the true-height electron profiles deduced from ionograms taken at Boulder at 1600 and 1700 m.s.t. The available orbital elements have been adjusted to give the correct time of closest approach to better than 1 sec. In the ray tracing for this particular pass, the 48-term expansion of the magnetic field was used, in order to obtain maximum accuracy. The derivation of $C(t)$ was made as described in section 3 above, using the full ray tracing program. The ray tracing program permits an estimate of the total number, D, of half rotations existing at the time of the first Faraday rotation null; in order to further refine this value, a family of curves differing only in value of D was plotted. The derivation of the most probable value of D was made as follows.

In a fixed, spherically stratified ionosphere, the rate of change of the subsatellite electron content is

$$\frac{dI}{dt} = N_s \cdot \frac{dh}{dt},$$

where N_s is the electron number density at the satellite, and dh/dt is the rate of increase of height of the satellite. For given orbital elements, dh/dt is known and it is therefore possible to compute observational values of N_s for each value of D as a function of time throughout the satellite passage. When this is done, very large differences in N_s are obtained for different values of D, particularly near the quasi-transverse region of the sky. Thus, the values of N_s deduced for the last nine points of each D curve of figure 4 vary from negative values ($D=82$), to approximately zero ($D=83$), a value approximately equal to the expected N_s ($D=84$) and values roughly $1\frac{1}{2}$ and 2 times greater than expected ($D=85$ and 86 respectively). For this reason, one can state the value of D almost certainly lies in the range 83 to 85.

Taking a value of $D=84$, one obtains a value of $I=3.62\times10^{12}$ electrons per square centimeter column at the time of closest approach. The corresponding value of I deduced from the model (itself based on the ionosonde data) is 3.46×10^{12}, a discrepancy of 4.5 percent. The difference between I and I_M is roughly constant throughout the satellite pass, and can very readily be explained by the effect of errors in D, in the magnetic field, in the ionospheric model used, in the assumed height of the satellite, and/or in the neglect of satellite spin and ionospheric gradients.

One striking feature of the $I(t)$ values derived in this way is that the values do not lie on a smooth curve. Figures 5, 6, and 7 are plots of $\Delta I/I_L$ for three passes at heights of the order 230, 300, and 1250 km. As indicated in section 3, the irregularities are relative to running smoothed values of I, the
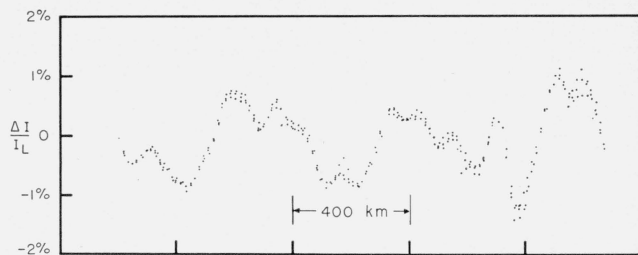

FIGURE 5. *Irregularities in subsatellite ionospheric electron content, 1620 m.s.t., August 1, 1958, as observed on three independent equipments.*
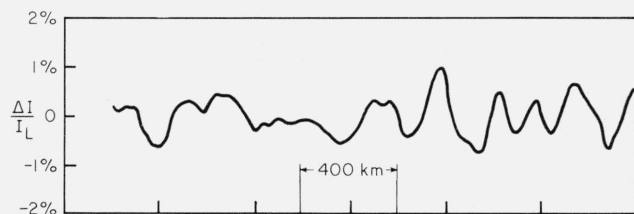

FIGURE 6. *Irregularities in subsatellite ionospheric electron content, 0830 m.s.t., September 4, 1958.*
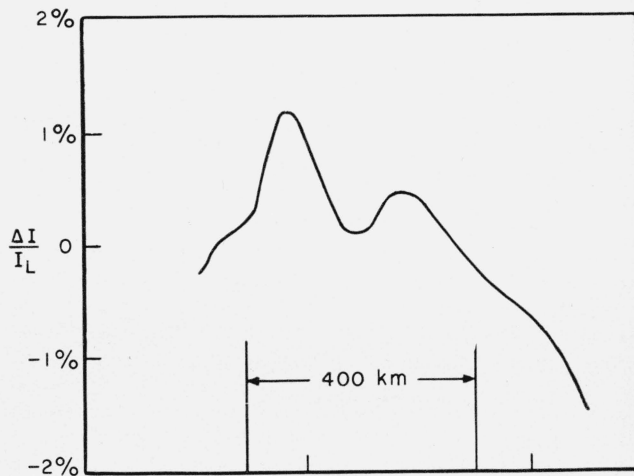

FIGURE 7. *Irregularities in subsatellite ionospheric electron content, 1330 m.s.t., September 17, 1958.*

smoothing interval being adjusted for each pass to correspond to a 400-km running average at a height of 300 km.

Figure 5 is a mass plot of the values of $\Delta I/I_L$ resulting from the Faraday-rotation data taken simultaneously on three separate receivers coupled respectively to dipoles oriented N–S, E–W, and from NW to SE. The excellent agreement of the points strongly indicates that the irregularities in polarization rotation which gave rise to the irregularities in I are real and are not due to such effects as scaling errors, or satellite signals scattered by objects near the antennas.

Figures 6 and 7 are of interest in that the irregularities observed on two very different passes are similar in fractional deviation and in size. Figure 6 refers to a south-north pass at a height of 290 km on September 4, 1958, and figure 7 to a 1,250-km height, north-south pass on September 17, 1958. Also of interest is that a pass observed by Dr. Owen Garriott of Stanford University, and analyzed by our method, shows similar irregularities. This fact would appear to confirm that the phenomenon is real, and is not introduced by our equipment.

5. Discussion of Accuracy

The accuracy with which I, the total electron content per unit column to the height of the satellite, can be derived by the Faraday-rotation method depends upon many factors, although the accuracy of measurement of the irregularities in I is primarily limited by scaling errors. The magnitude of typical scaling errors is illustrated by the scatter of the points in figure 5. It is quite small compared to the magnitude of the ionospheric irregularities themselves.

In contrast to angle-of-arrival studies, the Faraday method does not require extremely accurate knowledge of the position of the satellite. For example, the low pass illustrated in figure 4 was analyzed independently with two different sets of orbital elements. These were Vanguard predictions issued respectively 1 day before and 6 days after the passage of the satellite. They gave apparent minimum zenith angles at Boulder of 1° and 10°, and heights at closest approach of 236 and 232 km, respectively. In each case an arbitrary time correction of nearly a minute was needed to obtain agreement with the observed time of closest approach at Boulder. Despite the definite disagreement between the two sets of elements, the resulting values of I differed by less than 3 percent in all parts of the sky except the far north, where transverse propagation and the resulting critical dependence upon θ increased the difference to 15 percent.

It has already been mentioned that $C(t)$, and therefore $I(t)$, are nearly independent of the ionosphere model used in the analysis. The extent to which this is true is illustrated in figure 8, where the analysis has been carried through for two ionospheric models differing by a factor of two in electron density at each height. The resulting values of I differ by only 3 percent near the zenith, and by less than 10 percent at the ends of the recorded passage. As pointed out earlier, this fortunate behavior results from the fact that a change in I_M causes a nearly proportional change in Φ_M.

In figure 9 we demonstrate, using the passage of August 1, 1958, the importance of utilizing an accurate model for the geomagnetic field. The lowest curve shows the $I(t)$ which results from the use of the complete spherical-harmonic expansion of the field. The Gaussian coefficients used here are those published by Finch and Leaton [6]. In the absence of better knowledge of local geomagnetic anomalies,
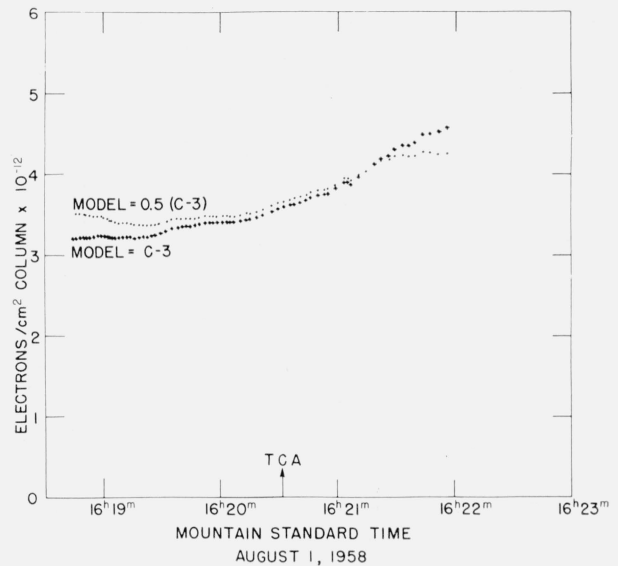

FIGURE 8. *Effect of different ionospheric models upon the derived subsatellite electron contents.*
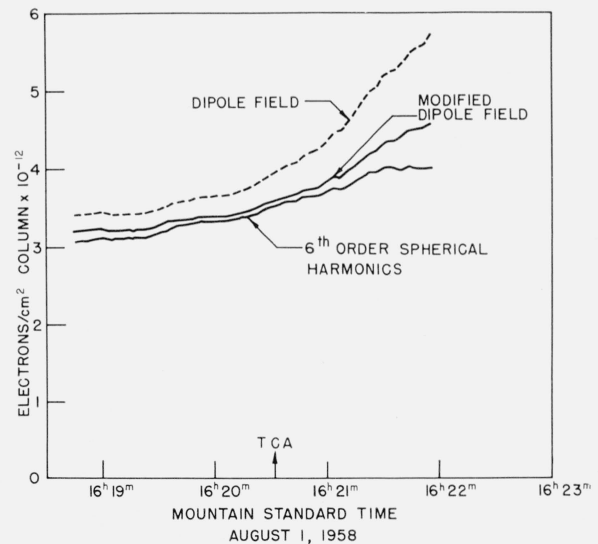
TCA = time of closest approach.

FIGURE 9. *Effect of different magnetic field models upon the derived subsatellite electron contents.*

including ionospheric currents, these coefficients represent the most accurate means we have available for estimating the geomagnetic field in the ionosphere. Figure 9 shows, for comparison, the result of using the standard centered dipole field, and the field of our "modified dipole." The modified dipole, adjusted to agree with the field at Boulder, removes about two-thirds of the total error caused by the standard centered dipole. Although an accurate magnetic model is necessary for estimating I, there is no such requirement for $\Delta I/I_L$. The modified

341

dipole or, if we except the transverse-propagation region near the northern horizon, even the standard dipole, would suffice for the study of the irregularities, since these are measured relative to the corresponding smoothed values.

Figure 10 illustrates the importance of including ionospheric refraction. Here the correct value of $I(t)$, including all refractive effects, is compared with another which neglects only refraction. This latter curve results from evaluation of $(2\pi/\lambda)\int_0^s (\mu_0-\mu_e)dl$ along the straight line joining the observer and the satellite. The complete Appleton-Hartree equation (excluding absorption) is used to evaluate μ_0 and μ_e, so the high-frequency approximation is avoided. As might be expected, neglect of refraction has no effect near the zenith but becomes serious near the horizon. At the northern horizon the effect is more pronounced than in the south, and it is of opposite sign. These differences result from the dependence of Faraday rotation upon $\cos \theta$, a factor which varies rapidly near the northern horizon.
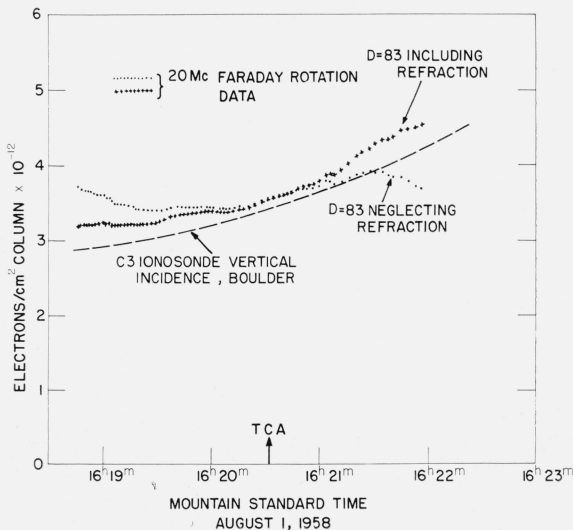

FIGURE 10. *Effect of including and omitting ionospheric refraction upon the derived subsatellite electron contents.*

The preceding discussion of accuracy assumes the existence of a spherically stratified ionosphere upon which irregular fluctuations may be superposed. Most of the remarks, particularly those concerning $\Delta I/I_L$ continue to be valid in the presence of systematic horizontal gradients, but there is one notable exception. The determination of D, and therefore of the absolute value of I, may be seriously influenced by such gradients. The implications of this effect are discussed in some detail in the next section.

6. Discussion and Interpretation of the Results

The discussion and interpretation of the results of analyses presented in earlier sections may be divided into (a) those dealing with the $I(t)$ values directly and (b) those dealing with the observed irregularities in $I(t)$.

6.1. Total Content

In view of the excellent agreement (for passes below F_{max}) between the derived $I(t)$ curves and those based on the corresponding ionosonde data, it may be claimed that the method is an accurate one and is capable of deriving the subsatellite electron content with relatively little uncertainty. It should be emphasized that the agreement shown in figure 4 is in no way exceptional and in fact is not as good as that obtained on at least one other pass. This excellent agreement may be taken as confirmation of the absence of any deep valley in the electron profiles between the E and the F regions, since the ionosonde profiles were derived assuming the absence of such valleys.

The main limitation to the accuracy is in the determination of D by the comparison of the slope of the model $I_M(t)$ and the derived $I(t)$ curves. This method is unfortunately sensitive to horizontal gradients in the electron content, and without prior knowledge of these gradients, it is apparent that significant errors could be made in selecting D. It should be realized that this difficulty applies also to the Faraday "fading-rate" method used by several authors [7, 8, 9].

Investigation of the change in dI/dt with D shows that a gradient of $\int_0^s N\,dh$ of 2 percent per 100 km along the pass would change the fading rate in the zenith, and hence the deduced value of $\int_0^s N\,dh$, by about 12 percent. Such gradients of $\int_0^s N\,dh$ must be expected, since even if they existed from pole to equator they would produce only a four-fold variation of total ionization. Over such distances the f_0F2 values typically vary by a factor of 3, implying, for low passes at least, a nine-fold variation in subsatellite electron content.

It should also be realized that, once D has been determined unambiguously, (for example, by simultaneous polarization measurements on two closely spaced frequencies) then the present method will become a sensitive one for investigating ionospheric gradients.

When the satellite is at a height well above the F-peak it is not possible to base the model ionosphere upon ionograms alone. Some assumption must be made concerning the shape of the upper portion of the electron density profile. As mentioned previously, the Chapman function is a convenient one to use. Its adjustable parameter, the electron scale height, must be given some arbitrary trial value. If there is reason to suspect that horizontal gradients may exist, it is, in addition, necessary to make some assumptions concerning them. Figure 11 illustrates the result of such an analysis of the record from a high satellite. The passage is the one shown in figure 2, from which it is obvious that some horizontal gradient should be included in the ionospheric model.
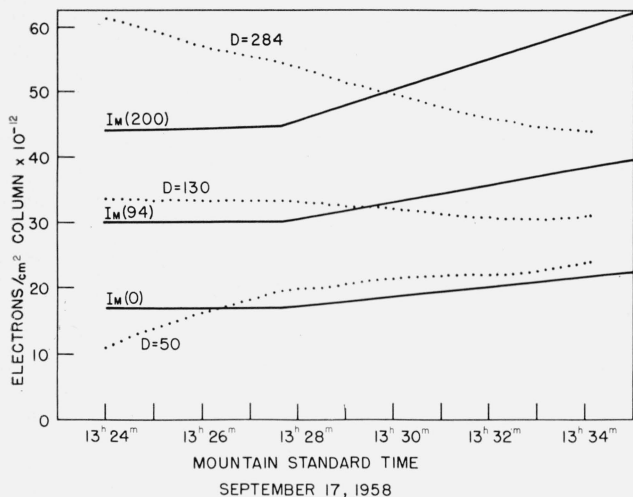
FIGURE 11. *Diagram illustrating the failure of the experimental data to fit the nonspherical ionosphere deduced from the f_0F2 data of figure 2.*

The simplest assumption is that the electron density varies with geographical position in the same way at all heights, and that f_0F2 can be used as an indicator of this variation. Acting upon this assumption we obtain for $I_M(t)$ the three curves shown in figure 11, for assumed values of electron scale height at Boulder of 200 km, 94 km, and zero. Zero scale height means, of course, that no electrons exist above the F-peak. The other three curves in the figure illustrate the resulting values of $I(t)$ for various values of D. The $I(t)$ curves shown were calculated with the ionospheric model which has an electron scale height of 200 km; the corresponding $I(t)$ curves for the other models are essentially the same.

A self-consistent picture demands that $I(t)$ agree with $I_M(t)$, and in this case it is obvious from figure 11 that no combination of scale height and D can produce such agreement. A change must be made in the ionospheric model, in such a sense as to reduce the rate of increase of subsatellite electron content with time. The f_0F2 values increased markedly along the track, and we have assumed an equivalent increase in electron density at all heights. Figure 11 shows that this assumption cannot be valid, and that the gradients must become relatively smaller above F-peak. In other words, the electron density at levels well above the F-peak must be less dependent upon geographical location than at F-peak.

The quantitative introduction of a new parameter, the rate of change of the horizontal gradient with height, is clearly unwarranted without additional data, though it may become possible to estimate this parameter after a large number of satellite records have been analyzed.

6.2. Irregularities

The irregularities in electron content shown in figures 5, 6, and 7 are of considerable interest, primarily because of their large size. These irregularities are, of course, derived from polarization-rotation data, and before discussing their significance one should first consider the validity of their derivation. As discussed in section 4, the excellent agreement between the values of $\Delta I/I$ derived from three completely independent observing systems precludes the possibility that the irregularities are introduced by scaling errors, or by scattered signals. One is therefore forced to conclude that the incident polarization vector did indeed rotate irregularly. Such irregular rotation could be introduced by an irregular motion of the satellite, by spatial or temporal irregularities in the magnetic field, and by spatial or temporal irregularities in the electron content of the ionosphere.

The effect of the rotation of the satellite upon the observed polarization is dependent upon the nature of the spin. A rapid, irregular tumble of the satellite can be ruled out, owing to the almost complete absence of external forces. The nonexistence of such an irregular motion was demonstrated by certain of the radio observations, which showed a regular sequence of nulls, superimposed upon the normal Faraday rotation, and occurring simultaneously on all polarizations. These nulls are interpreted as due to satellite rotation, the signal fading each time the rotating dipole was directed toward the observer. The time interval between these nulls was approximately 45 sec, implying a full rotation period of about 90 sec.

The effect of a regular rotation of the satellite upon the observed polarization is dependent upon the relationship between the orientations of the spin axis and the satellite antenna. In the case where the spin axis is parallel to the dipole, no effect will be observed. For orthogonal spin and dipole axes, the satellite spin will add to or subtract from the Faraday rotation, the correction reversing in sign as the observer passes through the plane of the rotating dipole. In the case where the angle between the spin and the dipole axes is neither zero nor 90°, it is convenient to differentiate between two conditions (a) where the observer lies inside the cone formed by the rotating antenna and (b) where the observer lies outside this cone. In general, there will be a transition between case (a) and case (b) as the satellite moves across the sky. In the region of sky defined by case (a), the satellite spin will add to or subtract from the Faraday rotation; for case (b) the observer will see the normal Faraday rotation plus a sinusoidal rocking of the polarization vector, of amplitude equal to the angle between the spin and the dipole axes, and period equal to the full rotation period of the satellite. In the former case the satellite rotation will introduce a cumulative error to the derived values of $I(t)$; for the latter case, quasi-sinusoidal oscillations of $I(t)$ of amplitude not greater than that introduced by a $\pm 90°$ shift in polarization rotation will be observed.

The derived irregularities of figures 5, 6, and 7 are not consistent with the suggestion that they are due to a regular satellite rotation. The irregularities are not sinusoidal, as would have been expected if they were due to satellite rotation. Instead of a period of about 90 sec, the irregularities

343

of figures 5 and 6 had periods of the order 40 to 50 sec, while those of figure 7 had periods considerably greater than 100 sec. Although the amplitude of the polarization flutter responsible for the irregularities of figures 5 and 6 was less than 90°, and was therefore consistent with the satellite-rotation hypothesis, the polarization irregularities observed during the pass of figure 7 were several times too large to be due to a regular satellite rotation. In addition, the amplitude of the irregularities in $\Delta I/I_L$, if due to satellite rotation, would be inversely proportional to the total amount of Faraday rotation existing at any time. A four- or five-fold increase in the amplitude of the irregularities should therefore have been observed in figures 5 and 6 as the satellite moved from the southern horizon toward the quasi-transverse region; no such increase was observed. For these reasons, therefore, we conclude that the observed irregularities in polarization angle were not due to satellite rotation.

Owing to the magnitude of the phenomenon (roughly ±1 percent variation in total amount of polarization rotation), it does not seem possible that the observed irregularities could be due to temporal variations in the magnetic field or in the subsatellite electron content. For the same reason, it does not seem likely that they were due to spatial irregularities in the magnetic field. For this to be the case, the longitudinal component of the geomagnetic field would have had to vary irregularly, relative to our magnetic model, by roughly ±1 percent in distances of the order of 300 km. Since maps of the surface magnetic field do not show such irregularities in the central United States, it would seem unlikely that they could exist at heights of the order of several hundred kilometers. It should be possible to test for such an effect by correlating the irregularities deduced from similar satellite passes, but as yet this test has not been carried out due to lack of fully analyzed data.

There remains one final alternative explanation of the deduced irregularities. The values of $C(t)$ are derived by a curve-fitting process, and any irregular variations in the computed $C(t)$ relative to the true $C(t)$ for a spherically stratified ionosphere would show up as irregularities in the electron content. The curve fitting process has therefore been tested. An analytical expression having a variation similar to that of $C(t)$ was obtained, and nine points along it were determined. The curve fitting program was then used to derive the best fitting 5th degree polynomial for these nine points. The agreement between values derived from the polynomial expression and from the original analytical expression was always within 0.1 percent, indicating that irregularities in the derived $C(t)$ curves were not responsible for the deduced variations in electron content.

On the basis of the above discussion, we conclude that the observed irregularities are correctly ascribed to variations in subsatellite electron content. We proceed now to discuss the properties of the irregularities; to relate our observations to other observations of large scale ionospheric irregularities, and to evaluate the Faraday-rotation method relative to other methods for observing such irregularities.

The following points may be made concerning these observed variations in subsatellite electron content:

(a) They have spatial extent of the order 300 km.

(b) Their intensity is such as to modulate the subsatellite electron content by roughly ±1 percent.

(c) They are not limited to a narrow height range (otherwise the fractional deviation would be considerably less for high passes than for low passes).

(d) On occasion, at least, they were present for heights as low as 230 km.

(e) They are of common daytime occurrence at latitude 40 deg. The four passes analyzed in detail to date all show their presence; a cursory examination of the observational data suggests that they are present on many, perhaps even the majority, of the records.

(f) Their intensity apparently is limited to maximum values of fractional deviation of the order of a few percent.

(g) There is no evidence as yet concerning their diurnal, seasonal, or magnetic storm dependence, though information concerning these is presumably available in the numerous, as yet unanalyzed, records.

Several authors have given experimental evidence of the existence of large-scale ionospheric irregularities. The observations of large-scale traveling disturbances in the ionosphere by Munro [10] suggest that the irregularities he observed are related to our own observations. From Australian studies of the virtual height of reflection of HF radio waves, he was able to deduce the frequent occurrence, during daytime hours, of traveling disturbances in the F region having dimensions up to several hundred kilometers. Bramley and Ross [11] using angle of arrival measurements, and Whitehead [12], and Jones et al. [13] from studies of the "bursts" in amplitude of ground-based transmissions have been able to confirm the existence of similar irregularities in the northern hemisphere. In general, however, quantitative data on the intensity of the irregularities is lacking, and it is therefore not possible to be certain that the irregularities we observe are of the Munro type; in particular, our observations to date do not permit the study of their motion.

Evidence of large-scale ionospheric irregularities has also been obtained by Vitkevitch and Kokurin [14, 15] using radio astronomical sources. These authors made measurements of the angle of arrival of radio waves from the celestial radio sources and found evidence of slow oscillations in the apparent source position relative to the true source position. They estimate that the irregularities in total electron content of the ionosphere have lateral dimensions of the order 300 to 400 km and intensity of up to several percent, in good agreement with our observations of the variations in subsatellite electron content.

Unlike the satellite data, the nonsatellite methods described above have not given more than rough

estimates of the irregularities in ionospheric electron content; and in particular the nonsatellite methods do not readily permit the rapid study of the irregularities over wide geographical areas from a single site. It is of interest to consider whether the other parameters of a satellite radio wave (amplitude, frequency, and direction of arrival) may also be useful in the study of these large-scale irregularities. Examination shows that these alternative methods are unlikely to prove effective. The recordings of amplitude "bursts" of satellite signals, analogous to the amplitude studies of Whitehead, and Jones et al., would not appear to be a suitable method, owing to difficulties arising from the very great variations of signal strength due to the changing satellite range, and the changing efficiencies of the polarized transmitting and receiving antennas along the (moving) line of sight. A second effect of the ionospheric irregularities is that they will modify the ionospheric distortion of the free-space Doppler curve of a moving satellite. Any Doppler method for studying these large-scale irregularities, however, suffers from the fact that the Doppler technique, which is based on the rate of change of phase, is inherently insensitive to weak, large-scale irregularities. Thus, it can be shown that the irregularities of figures 5 and 6 would produce an irregular distortion of the Doppler curve of magnitude about 0.1 cps and period roughly 40 sec. Since the rate of change of the 20-Mc signal is typically several cycles per second per second, it would be difficult to measure such distortions accurately. The angle-of-arrival measurements, which are based on the rate of change of phase across the ground, are also insensitive to weak, large-scale irregularities. In this case, it can be shown for the passes of figures 5 and 6 that the apparent position of the radio satellite would have varied irregularly, relative to its true position, by about 1 min of arc and with period about 40 sec. Such irregular refractions would be difficult to measure accurately in an object which is moving through as much as 100 min of arc per second. We conclude, therefore, that the Faraday rotation method is likely to prove the most advantageous satellite method for the study of large-scale ionospheric irregularities, and that it offers important advantages over previous nonsatellite methods.

7. Conclusions

The interpretation of the Faraday rotation of the plane of polarization of a satellite radio signal is at a relatively early stage of development. At the moment, the accuracy of measurements of subsatellite electron content is chiefly limited by uncertainty as to the total number of rotations. In the special case of a spherically stratified ionosphere, or an ionosphere whose gradients are known, this ambiguity can be largely removed. The technique is of considerable value for the study of large-scale irregularities, the interpretation of which is largely independent of ambiguities in the total number of rotations.

It may be expected that the Faraday-rotation methods will be developed still further. In particular, it would be extremely valuable if a satellite specifically designed for the purpose could be launched. Such a satellite should include two transmitters on rather closely spaced frequencies in order to permit resolution of the rotation ambiguity. The satellite should be designed so that the plane of polarization is unaffected by the rotation of the satellite, and some sensing mechanism should be available to define the orientation of the transmitting antenna. With such a satellite, it should be possible to measure the subsatellite electron content to an accuracy of about 1 percent, the accuracy being limited primarily by uncertainties in the magnetic field within the ionosphere. Furthermore, the sensitivity of the method to irregular variations in electron content could be considerably enhanced, particularly in the case of a satellite in a geostationary orbit. For such a satellite, it should be possible to observe variations in the polarization rotation of (say) 0.01 rotation; if this were done at frequency of 20 Mc, it should be possible to measure changes of $\Delta I/I_L$ as small as 1.0×10^{-4}.

8. References

[1] W. A. S. Murray and J. K. Hargreaves, Lunar radio echoes and the Faraday effect in the ionosphere, Nature **173,** 944 (1954).

[2] F. B. Daniels, Electromagnetic propagation studies with a satellite vehicle, ch. 30, Scientific uses of earth satellites, ed. J. A. Van Allen (Univ. of Michigan Press, Ann Arbor, Mich., 1956).

[3] J. A. Ratcliffe, The magneto-ionic theory and its applications to the ionosphere (Cambridge Univ. Press, Cambridge, Mass., 1959).

[4] H. Bremmer, Terrestrial Radio Waves, p. 305 (Elsevier Publishing Co., New York, N.Y., 1949).

[5] R. L. Anderson, Tables of orthogonal polynomial values extended to $N=104$, Agr. Expt. Sta., Iowa State College, Ames, Iowa, Research Bull. 297 (April 1942).

[6] H. F. Finch and B. R. Leaton, The earth's main magnetic field—epoch 1955.0—Monthly Notices, Geophys. Suppl. **7,** 314 (1957).

[7] F. B. Daniels and S. J. Bauer, The ionospheric Faraday effect and its application, J. Franklin Inst. **267,** 187 (1959).

[8] T. G. Hame and E. M. Kennaugh, Recordings of transmissions from the satellite 1958 Δ_2 at the Antenna Laboratory, Ohio State Univ., Proc. IRE **47,** 991 (1959).

[9] R. Parthasarathy and G. C. Reid, Signal strength recordings of the satellite 1958 δ 2 (Sputnik III) at College, Alaska, Proc. IRE **47,** 78 (1959).

[10] G. H. Munro, Travelling disturbances in the ionosphere, Proc. Roy. Soc. **A202,** 208 (1950).

[11] E. N. Bramley and W. Ross, Measurements of the direction of arrival of short radio waves reflected at the ionosphere, Proc. Roy. Soc. **A207,** 251 (1951).

[12] J. D. Whitehead, The focussing of short radio waves reflected from the ionosphere, J. Atmospheric and Terrest. Phys. **9,** 269 (1956).

[13] J. L. Jones, B. Landmark, and C. S. G. K. Getty, Movements of ionospheric irregularities observed simultaneously by different methods. J. Atmospheric and Terrest. Phys. **10,** 296 (1957).

[14] V. V. Vitkevich, Investigation of ionospheric inhomogeneities by radio-astronomical methods, Radiotekhnika i elektronika **3,** 478 (1958) (in Russian).

[15] V. V. Vitkevich and Y. L. Kokurin, Irregular radio-wave refraction and extensive discontinuities in the ionosphere, Radiotekhnika i elektronika **2,** 826 (1957) (in Russian).

(Paper 64D4–6 6)