X-625-71-186
PREPRINT

THE P'(f) TO N(h) INVERSION PROBLEM IN IONOSPHERIC SOUNDINGS

JOHN E. JACKSON

MAY 1971

GSFC
GODDARD SPACE FLIGHT CENTER
GREENBELT MARYLAND

THE P'(f) TO N(h) INVERSION PROBLEM IN IONOSPHERIC SOUNDINGS

by

John E. Jackson
Laboratory for Planetary Atmospheres
NASA Goddard Space Flight Center
Greenbelt, Maryland

ABSTRACT

A general review is given of the inversion techniques which are used to derive the ionospheric electron density N as a function of altitude h from group path P' versus frequency f measurements obtained by vertical incidence ionospheric sounders. The paper discusses the medium under investigation, the experimental techniques used to obtain the P'(f) data, the theoretical considerations which lead to the integral equation relating P'(f) to N(h), and the assumptions made in the inversion process. The lamination inversion technique is then presented, with special attention given to mathematical difficulties arising from discontinuities in the P'(f) function, infinities in the integrand, and in some cases unknown integration limits. Methods outlined for minimizing the uncertainties due to discontinuities include the use of redundant information (i.e. the two distinct P'(f) functions available for a given N(h) profile) and the use of models based upon statistical data. Mathematical procedures are

discussed which increase significantly the efficiency and accuracy of the required numerical integrations. Also included are recent refinements in inversion techniques which require fewer assumptions than are made in routine inversion. A number of examples are given to illustrate some of the procedures mentioned and also to point out their limitations. The accuracy of the inversion technique is deduced by comparing the resulting N(h) profile with N(h) data obtained by simultaneous but independent observations.

THE P'(f) TO N(h) INVERSION PROBLEM IN IONOSPHERIC SOUNDINGS

by

J. E. JACKSON

INTRODUCTION

Ionospheric data obtained by the vertical incidence pulse sounding technique have been used extensively to derive profiles of electron density N versus altitude h. The calculation of N(h) profiles from the P'(f) sounder data requires the inversion of an integral equation of the form:

$$P'(f) = \int_{p_0}^{p_r(f)} n' \left[N(p),\ f,\ B(p),\ \Phi(p) \right] dp \qquad (1)$$

where

P' = apparent range

f = frequency

p = propagation path

p_0 = location of sounder

p_r = location of reflection point for frequency f

n' = group refractive index (see next paragraph)

N = electron density

B = amplitude of terrestrial magnetic field vector $\vec{B}$

Φ = angle between $\vec{B}$ and direction of propagation.

The fundamental formula for n' is:

$$n' = n + f\frac{\partial n}{\partial f} \qquad (2)$$

where n is the real part of the refractive index of the medium. The index n is given by the well-known Appleton-Hartree formula, namely,

$$n^2 = 1 - \frac{2X(1-X)}{2(1-X) - Y_T^2 \pm \sqrt{Y_T^4 + 4(1-X)^2 Y_L^2}} \quad (3)$$

where

$$X = (fN)^2/f^2 \quad (4)$$

where

$$fN = \text{plasma frequency} = 8.98 \times 10^{-3} \sqrt{N} \quad (5)$$

(fN in MHz; N in electrons/cc

$$Y = fH/f \quad (6)$$

where

$$fH = \text{gyrofrequency} = 2.8B \quad (7)$$

(fH in MHz; B in gausses)

$Y_T = Y \sin\Phi$

$Y_L = Y \cos\Phi$

$\pm$ = positive sign in front of square root is for the ordinary wave; negative sign is for the extraordinary wave.

The ordinary wave reflects when X = 1 (except when $\Phi = 0$) or equivalently (as seen from Eq. 4) when:

$$f = fN \quad (8)$$

The extraordinary wave reflects when X = 1 - Y (if Y < 1) and also when X = 1 + Y. In P'(f) terminology, however, only the echo corresponding to X = 1 - Y is called an extraordinary echo, and the echo corresponding to X = 1 + Y is called the Z echo. Using Eq. 4 and 6 to express the reflection conditions in terms of fN and fH yields the frequency fx of the extraordinary echo:

$$fx = \tfrac{1}{2}(fH + \sqrt{4fN^2 + fH^2}), \quad (9)$$

and the frequency fz of the Z echo:

Equation (1) was written in terms of a generalized propagation path p to indicate that the sounder data do not necessarily correspond to "vertical" incidence. The propagation path is, however, essentially a vertical path, if the electron density distribution is only a function of altitude (spherically stratified ionosphere). Most P'(f) to N(h) inversion techniques assume that stratification is essentially spherical within the ionospheric region from which the sounding echoes are received and that the propagation path is vertical. Except when otherwise indicated these assumptions are also made in the present paper, i.e. in Eq. (1) p is replaced by h and Φ is replaced by $(90^{o} - \theta)$ where θ is the magnetic dip angle.

The formulas given above (Eq. 1-10) provide all the information required to evaluate P'(f) for a given N(p), B(p) and Φ(p). It is seen that these calculations would yield three P'(f) functions since there are three conditions under which reflection can take place. The calculations would require an explicit expression for n' in terms of X, Y_T and Y_L, which is obtained by a straightforward substitution of Eq. (3) into Eq. (2). The resulting formula (see Ref 1 p. 971) is rather complicated and it is not required for the present discussion. The complexity of the n' function makes it impossible to evaluate analytically the integral

shown by Eq. (1), except in very special cases ($\Phi = 0$ and $\Phi = 90$ degrees). Numerical integration techniques have to be used (even for a very simple N(h) model), and special care must be taken near the reflection points where n' becomes infinite.

It should also be noted that B and Φ in Eq. (1) are not unknown quantities, since they are specified by the sounder location and can be calculated to a high order of accuracy from a terrestrial field model (Ref 2). Thus the only unknown in the inversion process is N(h), letting h = p.

The inversion of Eq. (1) is, to a large extent, performed using the same mathematical techniques as are required for the direct evaluation of Eq. (1). The difficulties encountered in the inversion process are due primarily to gaps and discontinuities in the P'(f) data. The redundancy of the data (availability of more than one P'(f) function) helps overcome these problems. But in many cases it is also necessary to have some knowledge of the gross features of the N(h) distribution.

A complete discussion of the inversion process must therefore include some basic information concerning the medium under investigation and the limitations of the P'(f) measuring technique. The characteristics of the ionosphere and of the sounder technique are such that two sounders (one on the ground and the other in an earth

orbiting satellite) are required to permit a complete determination of the N(h) distribution at a given location. The present paper will discuss the P'(f) to N(h) inversion problems associated with both types of soundings.

THE MEDIUM UNDER INVESTIGATION

Since the ionosphere is produced primarily by ionizing radiation from the sun (offset by subsequent recombination and/or transport of the electron-ion gas thus created), the N(h) distribution depends upon the level of solar activity and experiences diurnal, seasonal, and latitudinal variations. The terrestrial magnetic field influences not only the n' function, but also the N(h) distribution. Both effects are due to the field control (through the $\vec{V} \times \vec{B}$ mechanism) upon the motion (transport or radiowave induced oscillations) of the free electrons. Since the magnetic axis is not aligned with the rotation axis of the earth, the magnetic control produces longitudinal variations in the N(h) distribution (in addition to variations due to local time differences).

Although the variability of the medium can introduce some difficulties (or limitations) in the N(h) calculations, certain features of the daytime and nighttime N(h) distributions are sufficiently regular - at least at midlatitudes - to provide a basis for the conventional division of the ionosphere into

D, E, F1 and F2 regions. The boundaries between these regions have never been precisely defined, but it is usually understood that the D region is below 90 km, the E region between 90 and 140 km, and F1 region between 140 and 180 km, and the F2 region above 180 km. The typical features of the midlatitude N(h) distribution can be seen on Fig. 1, which shows a daytime and a nighttime profile measured over Wallops Island, Virginia. An important characteristic of the distributions is the fact that the electron density becomes significant at an altitude which is about 80 km in the daytime and about 150 km at night (points Q on Fig. 1) and increases almost monotonically up to the height of maximum density (hmaxF2, shown on Fig. 1 as point M) which occurs typically at altitudes between 200 and 400 km. The major exception to this rule occurs in daytime in the region between 110 and 140 km (E valley). Above hmaxF2 (topside ionosphere) the electron density decreases monotonically and very nearly in exponential fashion.

EXPERIMENTAL DETERMINATION OF THE P'(f) FUNCTION

The ionospheric sounder, or ionosonde, operates on principles similar to those of radar, and it provides echoes from the ionosphere over a wide range of operating frequencies. The reflection conditions are such that 1) the ionosonde must utilize frequencies lower than

those normally used in radar and 2) a large number of sounding frequencies are required in order to investigate the plasma and obtain its electron density distribution. This last requirement is best met with the swept-frequency system, although single or multiple-fixed frequency sounders are also used for special applications. Ionospheric soundings are typically conducted at frequencies between 0.1 and 20 MHz, using 100 microsecond pulses and a 20-to-60 pulse-per-second repetition rate. A complete frequency sweep takes typically 15 to 30 seconds.

In the widely used swept-frequency system the received echoes are customarily displayed in the "ionogram" format, in which the echo round-trip time Δt is displayed (in the vertical axis direction) as a function of the sounder frequency f. The quantity $(\Delta t)c/2$ (where c = velocity of light in vacuo), which would be the target distance for a conventional radar, is called the apparent range; it represents a distance which is greater than the distance to the echoing region because the sounding signal is retarded by the ionosphere. The apparent range P' at a frequency f_r is therefore:

$$P'(f_r) = \frac{c}{2} (\Delta t_r) \tag{11}$$

where

$$\Delta t_r = 2 \int_{p_0}^{p_r} \frac{dp}{V_G} \tag{12}$$

$V_G = V_G(f,N,B,\Phi)$ = the velocity of the sounding signal,

p_r = a reflection point for the frequency f_r and the integration is along the path taken by the wave.

Combining Eqs. (11) and (12) gives

$$P'(f_r) = \int_{p_0}^{p_r} \frac{c}{V_G}\, dp. \tag{13}$$

In standard magneto-ionic nomenclature, the quantity c/V_G is called the group refractive index n'. Equation (13) then becomes

$$P' = \int_{p_0}^{p_r} n'dp. \tag{14}$$

as given in Eq. (1). The vertical (round-trip) axis of the ionogram is calibrated in terms of the apparent range P' and the basic measurement provided by the ionogram is the apparent range P' as a function of frequency.

CHARACTERISTICS OF TYPICAL P'(f) CURVES

The three reflection conditions (Eqs. 8, 9 and 10) give rise to the ordinary, extraordinary, and Z traces on ionograms. Conventional soundings from the ground rarely yield usable Z traces, but they normally provide both ordinary and extraordinary traces. Each of these P'(f) traces can be used (with appropriate inversion techniques) to derive N(h) profiles. The ordinary trace is generally the most useful one on ground-based soundings and the extraordinary trace is the most useful one on topside soundings (see p. 963 of Ref. 1).

In both cases (topside and ground-based soundings), the presence of at least one trace in addition to the one used for primary data analysis provides redundant information that can be used to check assumptions made in the analysis (i.e. in the inversion of Eq. (1)) and to minimize uncertainties due to gaps in the primary data. Further redundancy of information is also available on topside ionograms, as the result of resonance phenomena which occur at fN, fH, and at the upper hybrid frequency fT, which is given by:

$$fT = \sqrt{fN^2 + fH^2} \qquad (15)$$

These resonances permit the calculation of N and B at the satellite, which is particularly helpful when the propagation traces are not clearly defined at the satellite. Since N and B (consequently fN and fH) both decrease monotonically with altitude in the topside ionosphere, the P'(f) function is, in principle, known for both ordinary and extraordinary traces for the complete topside profile below the satellite (portion AM of profiles on Fig. 1). On ground-based soundings, however, the P'(f) function is not defined for the portion QR of the profiles on Fig. 1 (where R corresponds to the minimum density for which echoes are obtained by the sounder), and for the portion ST of the daytime profiles on Fig. 1.

The ordinary ray P'(f) functions corresponding to N(h) profiles of Fig. 1 are shown on Figs. 2 and 3 for

topside and for ground-based soundings. To permit a direct comparison between apparent heights P' and the actual reflection heights, the N(h) profiles have been redrawn with the density expressed in terms of plasma frequency (using Eq. 5).

It is seen that topside and ground-based soundings terminate at hmaxF2, which follows from the monotonic variation of fN and fx with electron density. Actually, the soundings do not quite reach hmaxF2; the topside soundings stop typically at a point M' (15 to 30 km above point M of Fig. 1) and ground-based soundings stop typically at a point M" (10 to 20 km below M). The region M'M" can be derived from simultaneous topside and ground-based soundings (Ref.3) or by extrapolation of either type of soundings. This extrapolation usually assumes that the N(h) distribution can be represented by a Chapman function in the vicinity of hmaxF2, i.e. N(h) is assumed to be given by:

$$N = N_{max} \exp\tfrac{1}{2}(1-z-e^{-z}) \qquad (16)$$

where

$$z = (h-hmax)/H$$

and H = scale height.

The unknown parameters in Eq. (16) are determined from the known portion of the N(h) profile near hmaxF2.

There is one additional source of information which can be used to check the total profile obtained

by simultaneous topside and ground-based soundings. Topside ionograms frequently display echoes reflected from the ground (ground trace or ground echoes). These echoes, which are for frequencies greater than the maximum frequency reflected from the ionosphere, exhibit a delay which is determined by the total profile (and by the frequency of the sounder signal). It is therefore possible to check the accuracy of the total profile by comparing the observed ground echoes with the ground echoes calculated from the total N(h) distribution (Ref. 3).

MATHEMATICAL TECHNIQUES USED FOR THE INVERSION OF EQ.(1)

The technique used for the inversion of Eq. (1) will be discussed for the case of vertical propagation, i.e. for the inversion of:

$$h'(f_r) = \int_{h_0}^{h_r} n' \left[N(h), f_r, B(h), \theta(h) \right] dh \qquad (17)$$

Equation (17) is written:

$$h'(f_r) = \sum_{i=0}^{i=k-1} \int_{h_i}^{h_{i+1}} n' dh \qquad (18)$$

where

$$h_k = h_r$$

$$n' = n' \left[N(h), f_r, B(h), \theta(h) \right]$$

For each interval $\Delta h = h_{i+1} - h_i$ (or lamination), the N(h) function is approximated by a simple analytic function. The number of laminations required can be minimized by choosing a function which approximates the N(h) distribution over large Δh increments. A

linear approximation is, however, adequate if small height increments are chosen and this simpler representation will be used to illustrate the procedure. The linear substitution:

$$h = h_i + a_{i+1}(N - N_i)$$

$$dh = a_{i+1}dN$$

gives for a typical lamination:

$$\int_{h_i}^{h_{i+1}} n'dh = a_{i+1} \int_{N_i}^{N_{i+1}} n'dN \qquad (19)$$

The integral corresponding to a given lamination is thus completely defined except for a constant a_{i+1}, which represents the value of dh/dN in the lamination. For all laminations, except the last one (h_{k-1},h_k), the function n' is finite* and the integral on the right hand side of Eq. (19) can be evaluated with great accuracy by numerical integration techniques, using, for example,a Gaussian quadrature formula with only three coefficients. The last lamination terminates at the reflectio point where n' is infinite. The integration in this case is based upon asymptotic expressions for n' near the reflection conditions, which are of the form:

$$n' = \alpha/\sqrt{1-\beta N}$$

where α and β are constants appropriate for the propagation

---

* There is one exception to this statement, and this occurs in the case of Z echoes on topside ionograms where n' becomes infinite at the start of the first lamination when $X = (1-Y^2)/(1-Y_L^2)$.

mode used. By letting $t^2=1-\beta N$, the last integral in Eq.(19) becomes:

$$\int_{N_i}^{N_{i+1}} n'dN = -\frac{2}{\beta} \int_{t_i}^{t_{i+1}} n't\, dt$$

At the reflection point, the integrand n't is finite since

$$n't = n' \sqrt{1-\beta N} = \alpha$$

Most of the techniques currently used for the inversion of Eq.(1) are based upon the above lamination concept. The laminations are implicitly specified by the choice of sampling frequencies, as illustrated by the following example based upon the ordinary h'(f) function for a topside sounding. Let $f_1, f_2, \ldots, f_n$ be the sampling frequencies and $h'_1, h'_2, \ldots, h'_n$ be the corresponding h' values. Let N_i be the density at which reflection occurs for the frequency f_i, i.e. (from Eq. 8 and 5) $N_i = 12400 f_i^2$, where N_i is in el/cc and f_i is in MHz. It is also assumed that N_1 is the density at the satellite, i.e. $h'_1 = 0$.

Using the linear approximation of Eq. (19), the virtual range values can be written:

$$\left.\begin{aligned} h'_2 &= a_2 \int_{N_1}^{N_2} n'(f_2)dN \\ h'_3 &= a_2 \int_{N_1}^{N_2} n'(f_3)dN + a_3 \int_{N_2}^{N_3} n'(f_3)dN \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ h'_n &= a_2 \int_{N_1}^{N_2} n'(f_n)dN + \ldots + a_n \int_{N_{n-1}}^{N_n} n'(f_n)dN \end{aligned}\right\} \quad (20)$$

Each of the integrals in Eqs.(20) can be evaluated using techniques previously discussed; thus, the unknown a_i terms can be readily computed.

For the extraordinary ray the procedure is slightly more complicated, because a knowledge of both f_i and the corresponding B_i at the height h_i are required to compute N_i (see Eq.9). The procedure in this case is to initiate the calculation of each new lamination with an estimated value of B and to refine this estimate by an iteration process (Ref. 1 p. 973-975 and Ref. 4).

The laminations commonly used for P'(f) to N(h) inversion are of the form:

$$h = h_i + a_{i+1}(\Psi - \Psi_i) + b_{i+1}(\Psi - \Psi_i)^2 \qquad (21)$$

where Ψ is either n, fN or ln N. When a parabolic lamination is used ($b_{i+1} \neq 0$), the additional lamination parameters are obtained by assuming that $dh/d\Psi$ is continuous at the lamination boundaries, i.e.

$$a_{i+1} + 2b_{i+1}(\Psi_{i+1} - \Psi_i) = a_{i+2} \qquad (22)$$

EXPERIMENTAL CONSTRAINTS PLACED UPON THE P'(f) FUNCTION

Some difficulties arise in the inversion of Eq.(1) due to the fact that the experimental data do not yield the P'(f) function for the complete altitude range of interest. These difficulties are encountered primarily with ground-based soundings due to regions QR and ST (see Fig. 1) for which the P'(f) functions are not available. In some cases on topside ionograms, the

extraordinary P'(f) function is not defined with sufficient accuracy at the starting point A. This usually arises when the density at the satellite is less than 200 electrons/cc. Some of the techniques developed for overcoming the above difficulties are discussed in the following sections.

UNKNOWN START ON TOPSIDE IONOGRAMS

If the density N_1 at the satellite altitude h_1 is unknown, it can be determined by the following method. The equations of the first two laminations are assumed to be linear in ln N, i.e.:

$$h = h_1 + a_2 \ln(N/N_1) \qquad \text{for } h_2 \leq h \leq h_1$$

and $$h = h_2 + a_3 \ln(N/N_2) \qquad \text{for } h_3 \leq h \leq h_2$$

It is then assumed that the correct value of N_1 will yield $a_2=a_3$. Various values of N_1 are tried (by an iteration process) until a_2 and a_3 agree within one-tenth of one percent. This slope matching technique has been tested extensively with ionograms where N_1 could be accurately measured by conventional scaling techniques. It was found that, for ionograms where N_1 is well defined, conventional and slope matching techniques have comparable accuracy. Results obtained with the slope matching technique have cast considerable doubt upon the determination of N_1 using the method introduced by Hagg (Ref. 5) which is based upon topside sounder plasma resonance observations under conditions

of low N_1. This method yields values of N_1 which are typically one-third of the slope matching technique values. Furthermore, the Hagg beat values (so-called because they are determined from the observed beat frequency between the fH and fT resonances) yield rather questionable N(h) profiles (see Fig. 4). Lockwood (companion paper) and Colin (private communication) have also used the slope matching technique (or similar procedures) to compute N_1 and they have reached similar conclusions concerning the Hagg beat. The low N_1 Hagg beat values have been attributed to plasma wave dispersion effects associated with the fH and fT resonances (Ref. 6).

UNKNOWN STARTS ON GROUND-BASED SOUNDINGS

The P'(f) function for a ground-based sounding can be written:

$$\int_0^{h_k} n'dh = h_Q + \int_{h_Q}^{h_R} n'dh + \int_{h_R}^{h_k} n'dh$$

where the letters Q and R refer to Fig. 1, and where h_k refers to a reflection point above point R. The altitude h_Q is the altitude below which n' can be considered equal to unity for the frequencies of interest. The group index n' cannot be considered equal to unity in the region QR even though this region yields no echoes on conventional ionograms. The procedures used

are different on daytime and on nighttime ionograms and these two cases will be discussed separately.

(a) daytime ionograms

One solution (See Ref. 7 p. 120-122), which is similar to the slope matching technique, is to assume that the region QR is a simple analytic extension of the profile above point R. Another approach, based upon rocket results (Fig. 5), is to assume that the density is 1000 el/cc at 80 km, and that it increases exponentially up to point R. This gives fairly good results with midlatitude, midday ionograms.

(b) nighttime ionograms

Analytic extensions of the profile above point R could also be assumed for the region QR, but more reliable results can usually be obtained by making use of the ordinary and extraordinary traces which are of comparable quality on nighttime ionograms. The procedure is to assume an arbitrary density value at Q and to compute from the ordinary trace the corresponding N(h) profile. The extraordinary P'(f) function corresponding to this profile is then computed and compared to the observed extraordinary P'(f) function at m values of sounding frequencies. The standard deviation σ between the computed and the observed extraordinary P'(f) values is:

$$\sigma = \sqrt{\sum(d_i^2/m)} \qquad (23)$$

where d_i represents the vertical height differences at $f = f_i$. The process is repeated for other assumed densities at Q and the density at Q which yields the minimum value of σ is taken as the starting density. The quantity σ for typical nighttime midlatitude ionograms is shown in Fig. 6 as a function of the assumed plasma frequency at Q. A value of h_Q = 150 km was used for this calculation. Also indicated on Fig. 6 is the altitude at which the electron density reaches a value of 10^5 el/cc. Extensive tests of this technique have given a high degree of consistency in the 150 km starting value obtained at one sounder location (Wallops Island), namely

$$0.2 \text{ MHz} \leq (fN) \leq 0.4 \text{ MHz}.$$

Thus an empirical starting value of fN = 0.3 MHz could be used in cases when the two traces cannot be scaled with sufficient accuracy to permit the above determination of N_Q.

THE VALLEY PROBLEM

The valley ambiguity (unknown region ST) can in principle be substantially minimized by using a technique similar to the one described for the unknown start on nighttime ground-based soundings. In this case a model of the E valley is arbitrarily selected and the quantity σ (Eq. (23)) is computed for various valley depths. The desired valley is

the one which yields the minimum value of σ. The effect of the assumed valley upon the resulting N(h) profile is shown in Fig. 7, which is for a Wallops Island ionogram obtained on 7 March 1970 at 1519 EST. The shape of the valley (a triangular wedge on this semi-logarithmic plot) is one of many assumed by other authors (Refs. 8 and 9). It is physically unrealistic but convenient for computer programming. Furthermore the shape of the valley is too variable to justify the use of a more aesthetic representation. The purpose of the valley is to improve the accuracy of the profile above the valley (rather than to provide meaningful data within the valley). The profiles shown in Fig. 7 would normally be derived from the ordinary trace. These profiles and two additional ones for a 90% and a 70% valley respectively have been used to compute the corresponding family of extraordinary traces above Emax shown in Fig. 8. The solid points which represent the observed X-trace are seen to agree best with a valley of 80%.

The valley determination illustrated by Figs. 7 and 8 was possible because the X trace was well defined for frequencies immediately above the extraordinary ray E-region critical frequency fxE. This requirement is often not met and it is seen, by inspection of Fig. 8,

that the comparison would have been much less meaningful if the ionogram had not shown an X trace for frequencies less than say 4.5 MHz (particularly in view of the $\pm$5 km scaling uncertainty).

The valley calculations (when feasible) yield fairly consistent results, and these results can be used to estimate the valley depth in cases when the extraordinary trace cannot be used for this purpose. For midlatitude midday conditions both valley calculations, as illustrated above, and rocket data (Refs. 10 and 11) show that the minimum E-valley density is typically 80 to 90% of the density at Emax (point S of Fig. 1).

PROBLEMS ARISING FROM NON-VERTICAL PROPAGATION

The routine procedure for the inversion of Eq. (1) is to assume that the P'(f) function represents virtual heights measured at vertical incidence. This assumption would, for B = 0, be equivalent to stating that the sounder is receiving echoes from a spherically stratified region of the ionosphere. This region for a ground-based sounder would be a vertical cylinder having a diameter of typically 10 km. For a topside sounder, which can move horizontally up to 200 km while a complete ionogram is obtained, the spherical stratification would have to hold over a vertical region whose projection

on the earth's surface would be roughly a 10 by 200 km rectangle. There is usually a change in altitude (and also some minor changes in B and θ) associated with the horizontal motion of the topside sounder. These effects can be taken into consideration (see companion paper by Lockwood) by assuming that the N(h) profile is a function of altitude only and that the successive soundings begin at a new reference altitude (and corresponding density).

The spherical stratification concept usually applies to the constant electron density contours and, for B = 0, this assumption would yield constant phase index contours. The presence of the terrestrial magnetic field introduces modifications in the constant phase index contours, and causes the ray paths to be deviated from the vertical mostly near the reflection points. Theoretical investigations of this problem (Colin, private communication) indicate that no serious errors are introduced in the virtual height calculations by assuming that the phase index contours are the same as the electron density contours. A more important effect of the magnetic field is the occasional presence of field-aligned irregularities which can act as wave guides and cause echoes (in the extraordinary mode) to be field-aligned. This can be recognized by an

experienced observer, and in this case the analysis (done for $\Phi = 0$) yields a field-aligned N(h) profile. From a sequence of these field-aligned profiles, a sequence of vertical profiles can be derived. The results of an analysis of this type are shown in Fig. 9.

Departures from spherical stratification can sometimes be inferred from a time sequence of ionograms. In the case of ground-based soundings it can be detected (when severe tilts are present) from the fact that multiple echoes (two-or-more round trips) do not give virtual heights in harmonic relationship to the first (one-round trip) echo virtual height. For example, if h' is the first echo virtual height, the second echo virtual height should be 2h' for spherical stratification. If the constant density contours make an angle γ with the horizontal, the second echo virtual height is $2h' \cos\gamma$. Departures from spherical stratification could also be inferred from a closely spaced network of ground-based sounders. Analysis techniques which can take into consideration these departures from spherical stratification are currently under investigation (see companion paper by McCulley). Routine inversion techniques, however, assume that the N(h) distribution corresponding to a given ionogram is a function of altitude only, and

that this N(h) distribution does not change significantly during the 10-to-30 seconds required to obtain an ionogram.

ACCURACY OF THE INVERSION TECHNIQUES

The accuracy of the mathematical techniques used in the inversion of Eq. (1) has been tested extensively (see Ref. 1 p. 967 to 970). These tests have shown that the integration techniques do not introduce significant errors in the inversion of Eq. (1). For example, on topside ionograms the error in altitude for a given density is typically less than one kilometer.

Some errors are introduced in the inversion of Eq. (1) for ground-based soundings due to the discontinuities in the P'(f) function at the starting point and at the E valley. These discontinuities introduce errors in altitude (up to 10 km) at the corresponding places on the N(h) profile, more specifically at the starting point and just above the E valley. The errors for the N(h) profile above the E valley decrease monotonically with altitude and near hmaxF2 the calculated N(h) profiles are affected at most by 1 or 2 km by the uncertainties at the lower altitudes.

It would therefore be expected that ionograms obtained simultaneously by a ground-based ionosonde

and by a topside sounder directly overhead would yield N(h) profiles which should agree near hmaxF2. Comparisons of this type have yielded disagreements (Ref. 3) which are several times greater than the errors attributable to scaling accuracy and to the mathematical techniques used for the inversion of Eq. (1). The results suggest that the error is roughly proportional to the length of the propagation path. Since the maximum propagation path (within the ionosphere) is usually several times greater for a topside sounding than for a ground-based sounding, one would attribute this discrepancy primarily to the calculated topside N(h) profile. On this basis the topside profile appears to be too low by about 3 to 5 percent of the distance d between the topside sounder and the reflection point. One test for the accuracy of the resulting total (or composite) profile is to calculate the corresponding total content $\int_0^{hs} Ndh$ (where hs is the satellite altitude) and to compare the results with an independent measurement of the total content. Comparisons of this type using data from GEOS-2 to compute $\int Ndh$ have shown consistently that the total content for the calculated composite profile was too low. By raising the topside profile by 3 to 4 percent of the distance d, the two measurements of $\int Ndh$ can be brought into agreement. In some cases, an additional

check was possible using the topside ground trace. This procedure is illustrated by Fig. 10 and 11. The composite profile (solid line) shown in Fig. 10 was obtained from simultaneous topside and ground-based observations. The corresponding ordinary and extraordinary ground traces are shown in Fig. 11 (solid line). The points (with indicated error bars) scaled from the observed ground traces are seen to be significantly lower than the computed ground traces. By raising the topside profile as shown by the dashed line in Fig. 10, the corresponding computed ground traces (dashed lines in Fig. 11) are brought in much closer agreement with the observed values.

The small but puzzling discrepancy between topside and ground-based sounding suggests that some cumulative errors arise in the observations due perhaps to irregularities in the ionosphere causing the propagation paths to deviate from the vertical direction. The discrepancy does not seem to be due to large-scale departures from spherical stratification, as could be inferred from the analysis of ionogram sequences.

Another possibility not discussed in the report is that cumulative errors arise in the analysis due to one of the many assumptions made in the magneto-ionic theory (cold plasma treatment of the ionosphere, WKB approxi-

mation, idealized reflection conditions, group velocity representation of the signal velocity, etc). Although these approximations have been accepted for several decades, a complete evaluation of their effects has not been made in the topside ionosphere.

SUMMARY

1)The P'(f) to N(h) inversion in ionospheric soundings is based upon the following assumptions:

a)spherically stratified medium (except in special techniques under development which make allowance for gradients),

b)medium which does not change during the sounding,

c)vertical propagation (except in more elaborate techniques used, for example, to deal with geomagnetic field-aligned propagation), and

d)simple ray theory of wave propagation in the ionosphere.

2)The most accurate inversion technique is based upon:

a)lamination concept and step-by-step solution,

b)lamination model optimized for representative N(h) distributions,

c)numerical integration with change of variable to keep the integrand finite, and

d)iteration (with topside extraordinary P'(f) function) to find the upper limit of integration.

3)Special problems arise, particularly with the P'(f) function obtained by ground-based soundings, when the P'(f) function is not defined over the complete altitude range of interest. This leads to:

   a)starting point problem, and

   b)valley problem (for daytime ground-based soundings).

4)The solutions to these special problems are based upon:

   a)redundancy of the data (ordinary and extraordinary P'(f) functions);

   b)estimated parameter values derived from statistical data accumulated from previous P'(f) to N(h) inversions, or from other techniques, such as in situ rocket measurements;

   c)(in some cases) extrapolations or "slope matching techniques".

5)The accuracy of the N(h) profiles derived by inversion of the P'(f) function is typically better than 5 percent of the length of the propagation path in the ionosphere. The errors in the N(h) profiles are apparently due to assumptions made in the analysis, since they are several times greater than the error attributable to the mathematical techniques used in the inversion process. Although the human element was not stressed in this paper, it is an important factor in the ultimate accuracy achieved with a given inversion technique.

The human element enters in the initial recognition and accurate scaling of the P'(f) function, in the evaluation of the validity of the assumptions made for a given experimental condition, and in the final acceptance (or rejection) of the results given by the inversion process.

REFERENCES

(1) Jackson, J.E., "The Reduction of Topside Ionograms to Electron-Density Profiles", Proceedings of the IEEE, June 1969, Vol. 57, No. 6, p. 960-976.

(2) Cain, J.C. and S. Cain, "Derivations of the International geomagnetic Reference Field (Oct. 1968)", NASA Goddard Space Flight Center, Greenbelt, Md., Rept. X-612-68-501, December 1968.

(3) Jackson, J.E., "Comparisons Between Topside and Ground-Based Soundings", Proceedings of the IEEE, June 1969, Vol. 57, No. 6, p. 976-985.

(4) Lockwood, G.E.K., "A Modified Iteration Technique for Use in Computing Electron Density Profiles from Topside Ionograms", Radio Science, Vol. 5, p. 575, 1970.

(5) Hagg, E.L., "Electron Densities of 8-100 Electron cm^{-3} Deduced from Alouette II High Latitude Ionograms", Can. J. Phys., Vol. 45, p. 27, 1967.

(6) Benson, R.F., "Plasma Wave Dispersion Effects Observed on the Alouette II nfH Resonances", NASA Goddard Space Flight Center Report X-621-71-214, May 1971.

(7) Jackson, J.E., "A New Method for Obtaining Electron Density Profiles from P'-f Records", J. Geophys. Res., Vol. 61, pp. 107-127, March 1956.

(8) Davies, K., and A.K. Saha, "Study of the 'Valley Problem' with a Ray Tracing Program", Electron Density Profiles in the Ionosphere and Exosphere, pp. 162-166 edited by B. Maehlum, New York, Pergamon Press, 1962.

(9) Mal'tseva, O.A., "Method of Correcting the N(h) Profiles of the F-region with Allowance for a 'Valley', Geomagnetism and Aeromony, Vol. IX, No. 5, pp. 753-755, 1969.

(10) Seddon, J.C. and J.E. Jackson, "Ionosphere Electron Densities and Differential Absorption", Annales de Geophysique, Tome 14, No. 4, pp. 456-463, Oct.-Dec. 1958.

(11) Bauer, S.J. and J.E. Jackson, "A Small Multi-Purpose Rocket Payload for Ionospheric Studies", Goddard Space Flight Center document X-615-63-95 presented at the Fourth International Space Sciences Symposium (COSPAR) in Warsaw, Poland, June 3-11, 1963.

(12) Bourdeau, R.E., A.C. Aikin, and J.L. Donley, "Lower Ionosphere at Solar Minimum", J. Geophys. Res., Vol. 71, pp. 737, February 1, 1966.

FIGURE CAPTIONS

Fig. 1 Vertical electron density distributions based upon simultaneous topside and ground-based observations at Wallops Island, Virginia (latitude 37.84° N, longitude 75.47° W, fH = 1.45 MHz, θ = 69°). The daytime curve is for June 18, 1968 at 1200 EST and the nighttime curve is for June 4, 1968 at 0245 EST. The two N(h) profiles corresponded to the same magnetic index (Kp = 3) and 10.7 cm solar flux (Sf = 144). The altitude difference between S and T was made about 15 km greater than observed in order to show the E valley more clearly.

Fig. 2 Vertical distribution of plasma frequency (dashed-curve) corresponding to the daytime N(h) profile of Fig. 1, and corresponding ordinary ray P'(f) functions (solid curves) for topside and ground-based soundings. The virtual range scale is one half the altitude scale to keep the ionograms from overlapping.

Fig. 3 Vertical distribution of plasma frequency (dashed-curve) corresponding to the nighttime N(h) profile of Fig. 1, and corresponding ordinary ray P'(f) functions (solid curves) for topside and ground-based soundings. The virtual range scale is one half the altitude scale to keep the ionograms from overlapping.

Fig. 4 Comparisons between two N(h) profiles derived from the same ionogram using the Hagg-beat start (dashed-curve) and the slope matching start (solid curve).

Fig. 5 Rocket results (Ref. 12) showing D and E regions. The lower part of the profile (for which the P'(f) function is not normally available) can be approximated by the indicated straight segments.

Fig. 6 The function σ versus plasma frequency at 150 km for typical midlatitude nighttime ionograms. The effect of the starting point upon the resulting N(h) profile is indicated by the dashed line which shows the altitude at which the density is 10^5 el/cc.

Fig. 7 Effect of assumed valley upon calculated N(h) profiles.

Fig. 8 Calculated X traces for N(h) profiles of Fig. 7 and for profiles corresponding to two additional valleys (not shown in Fig. 7). The observed X trace values are shown by the solid points.

Fig. 9 Electron density contours derived from topside field-aligned traces.

Fig. 10 Composite profile obtained by simultaneous topside and ground-based soundings at Wallops Island, Virginia. Solid curve is for calculated profile. Dashed curve is for revised profile based upon the ground trace calculations (see Fig. 11).

Fig. 11 Ground traces for calculated and revised profiles of Fig. 10. The observed ground trace data are the points with the indicated error bars.

Fig. 1

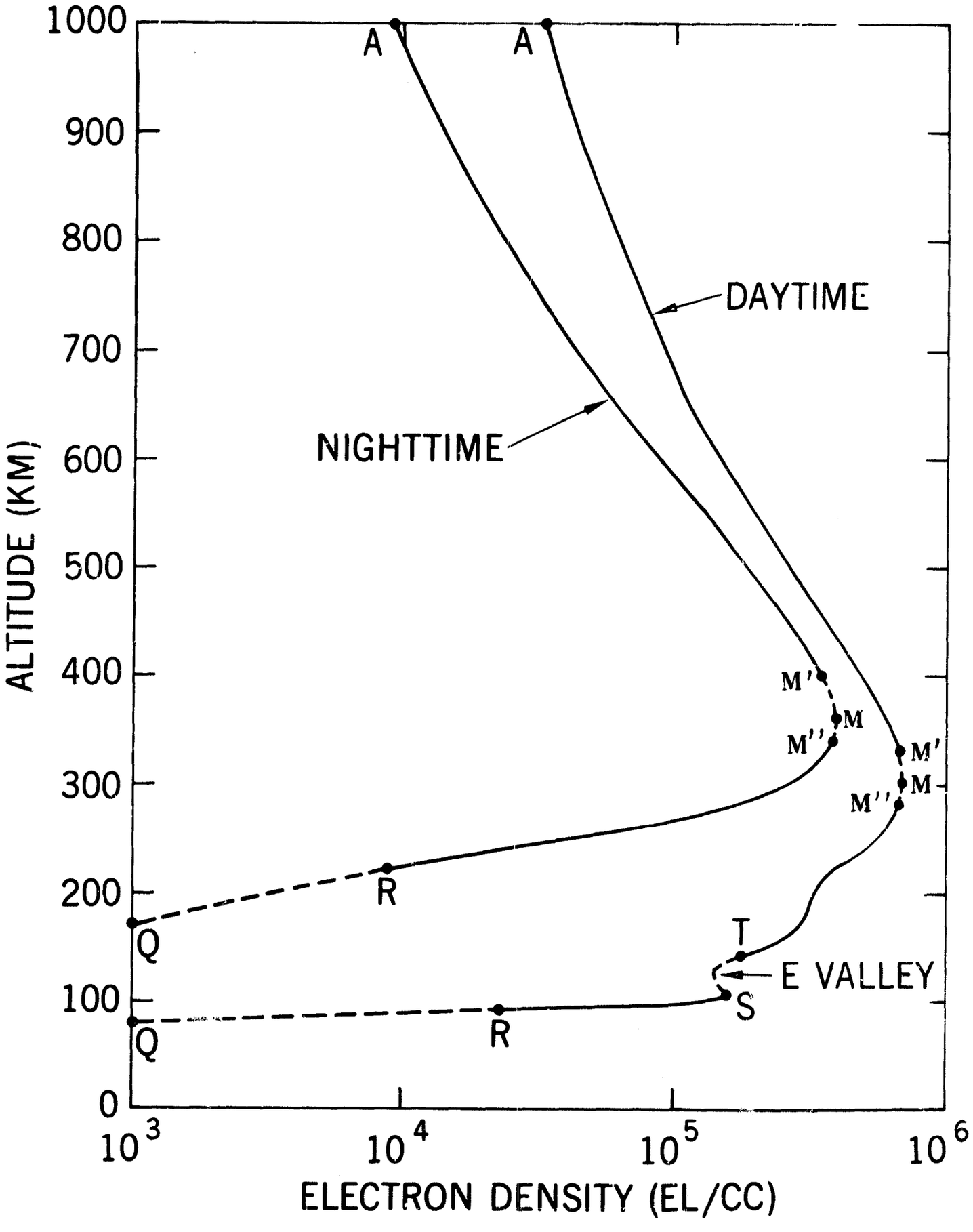

1000
900
800
700
600
500
400
300
200
100
0
ALTITUDE (KM)
A
A
DAYTIME
NIGHTTIME
M'
M
M''
M'
M
M''
R
Q
T
E VALLEY
S
Q
R
10^3
10^4
10^5
10^6
ELECTRON DENSITY (EL/CC)

Fig. 2

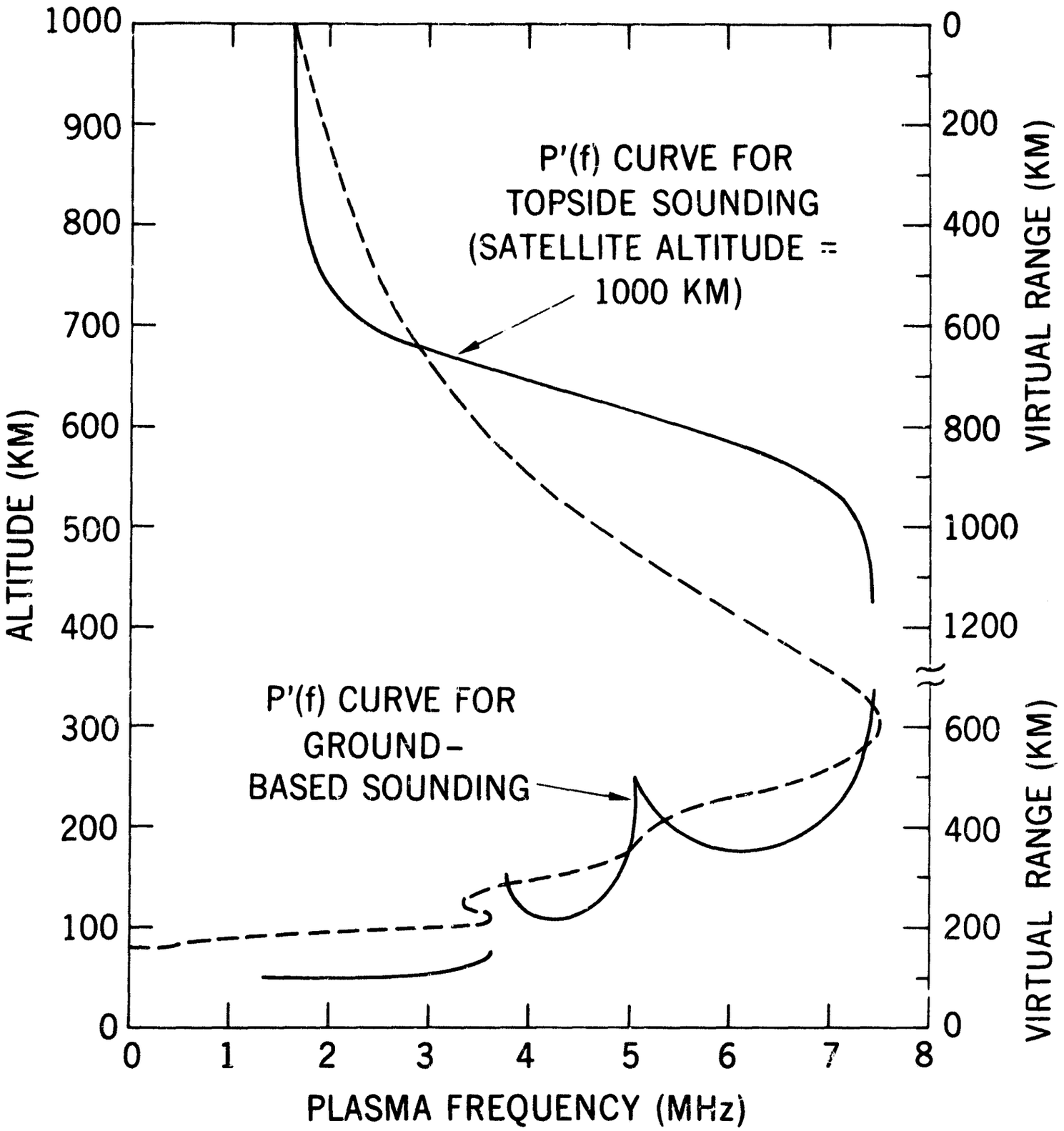

1000
900
800
700
600
500
400
300
200
100
0
ALTITUDE (KM)
0
200
400
600
800
1000
1200
600
400
200
0
VIRTUAL RANGE (KM)
VIRTUAL RANGE (KM)
P'(f) CURVE FOR
TOPSIDE SOUNDING
(SATELLITE ALTITUDE =
1000 KM)
P'(f) CURVE FOR
GROUND-
BASED SOUNDING
0
1
2
3
4
5
6
7
8
PLASMA FREQUENCY (MHz)

Fig. 3

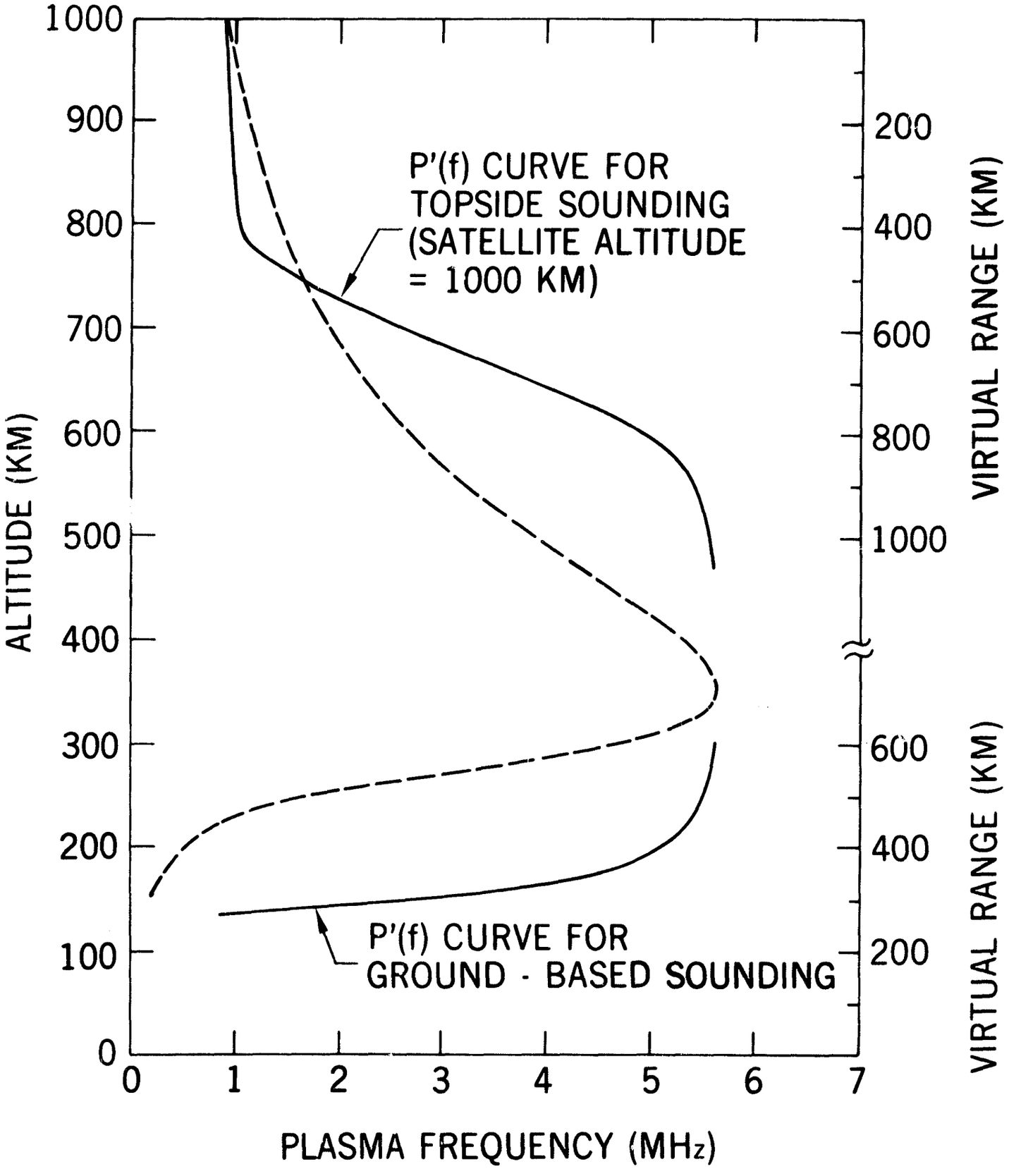

1000
900
800
700
600
500
400
300
200
100
0
ALTITUDE (KM)
P'(f) CURVE FOR
TOPSIDE SOUNDING
(SATELLITE ALTITUDE
= 1000 KM)
P'(f) CURVE FOR
GROUND - BASED SOUNDING
0
1
2
3
4
5
6
7
PLASMA FREQUENCY (MHz)
200
400
600
800
1000
VIRTUAL RANGE (KM)
600
400
200
VIRTUAL RANGE (KM)

Fig. 4

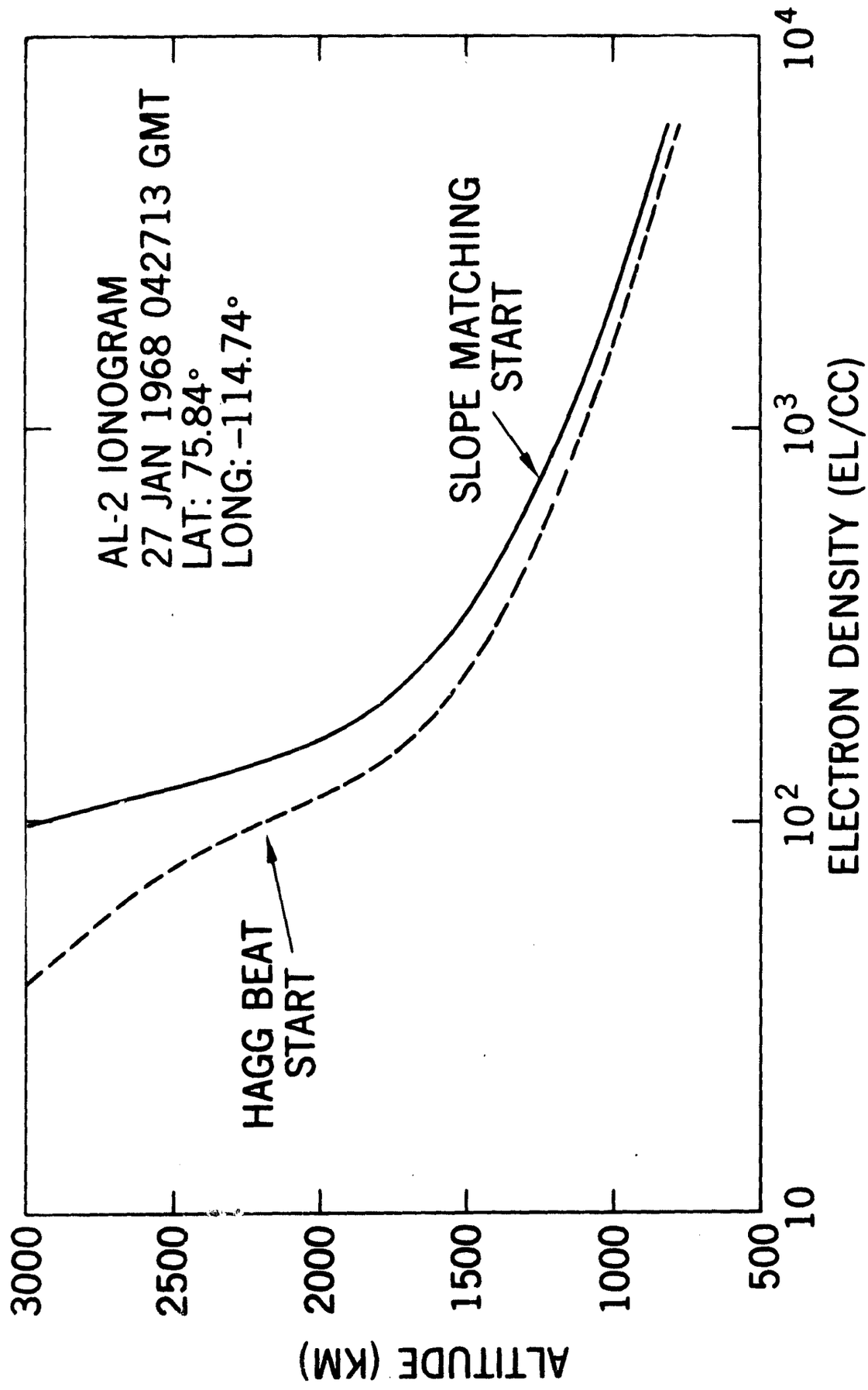

AL-2 IONOGRAM
27 JAN 1968 042713 GMT
LAT: 75.84°
LONG: -114.74°
SLOPE MATCHING START
HAGG BEAT START
ALTITUDE (KM)
3000
2500
2000
1500
1000
500
10
10^2
10^3
10^4
ELECTRON DENSITY (EL/CC)

Fig. 5

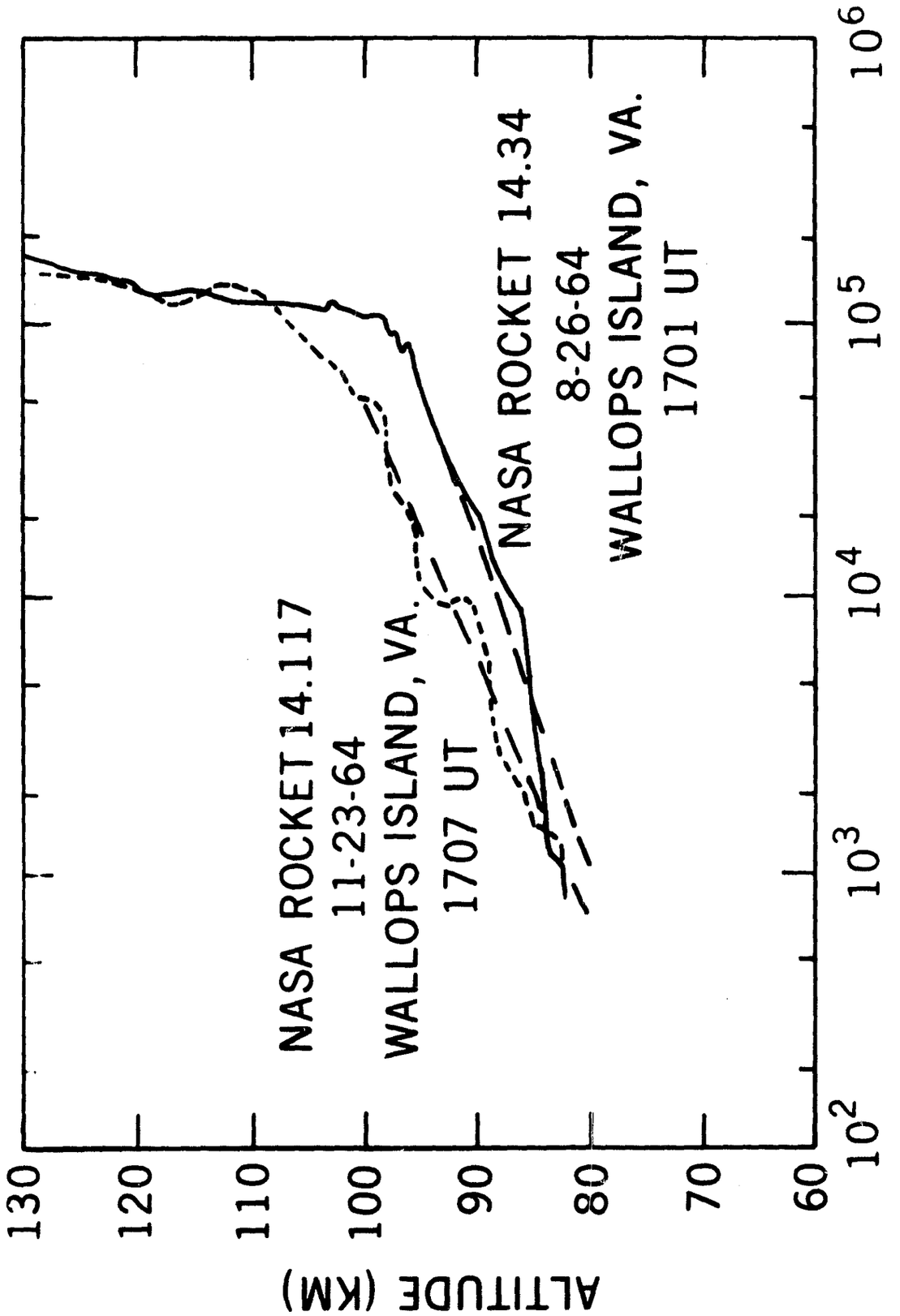
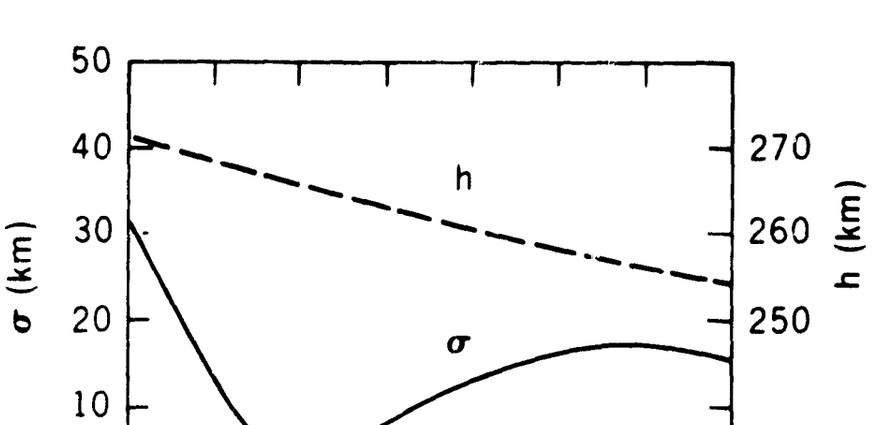
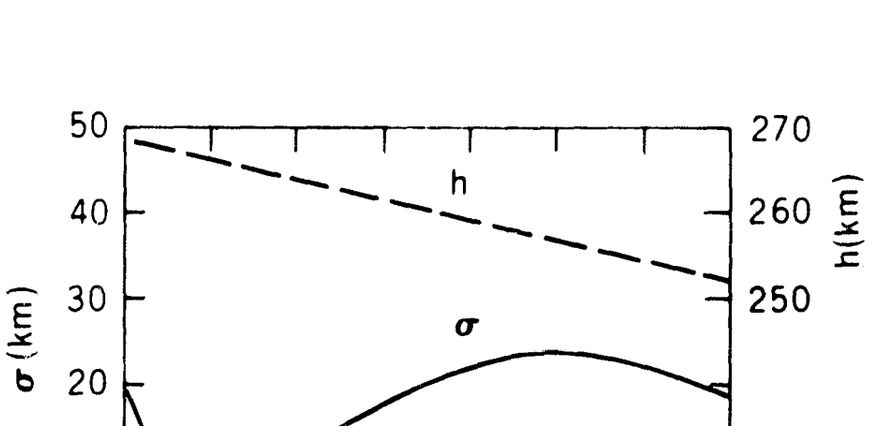

ALTITUDE (KM)
130
120
110
100
90
80
70
60
10^2
10^3
10^4
10^5
10^6
ELECTRON DENSITY EL/CC
NASA ROCKET 14.117
11-23-64
WALLOPS ISLAND, VA.
1707 UT
NASA ROCKET 14.34
8-26-64
WALLOPS ISLAND, VA.
1701 UT

Fig. 6



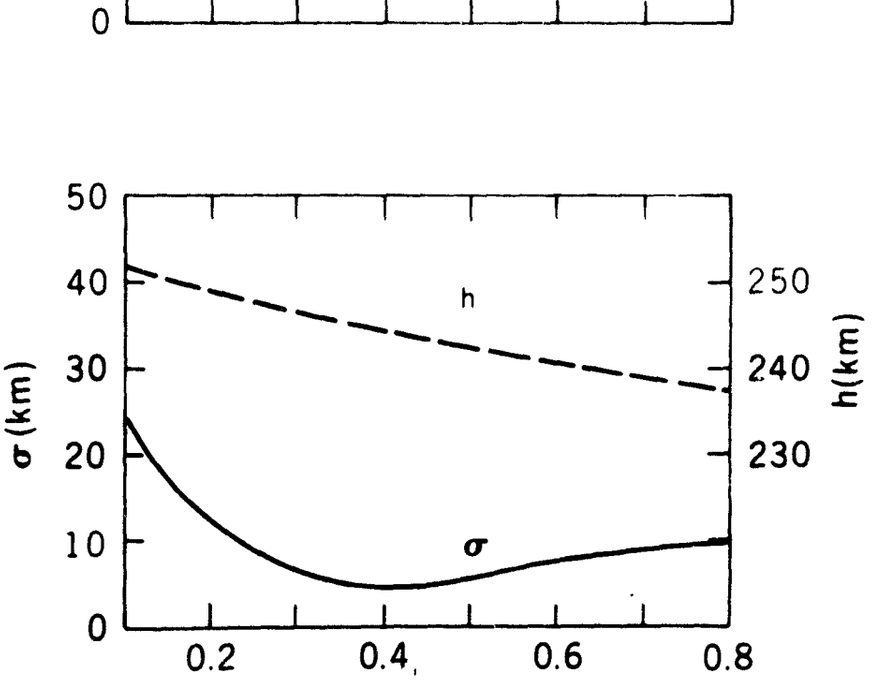
WALLOPS ISLAND
4 APRIL 1968
2145 EST

hmaxF2 = 380 km
NmaxF2 = 4.8×10^5



WALLOPS ISLAND
23 MAY 1968
2215 EST

hmaxF2 = 380 km
NmaxF2 = 7.6×10^5



WALLOPS ISLAND
28 MAY 1968
2145 EST

hmaxF2 = 370 km
NmaxF2 = 6.8×10^5

Fig. 7

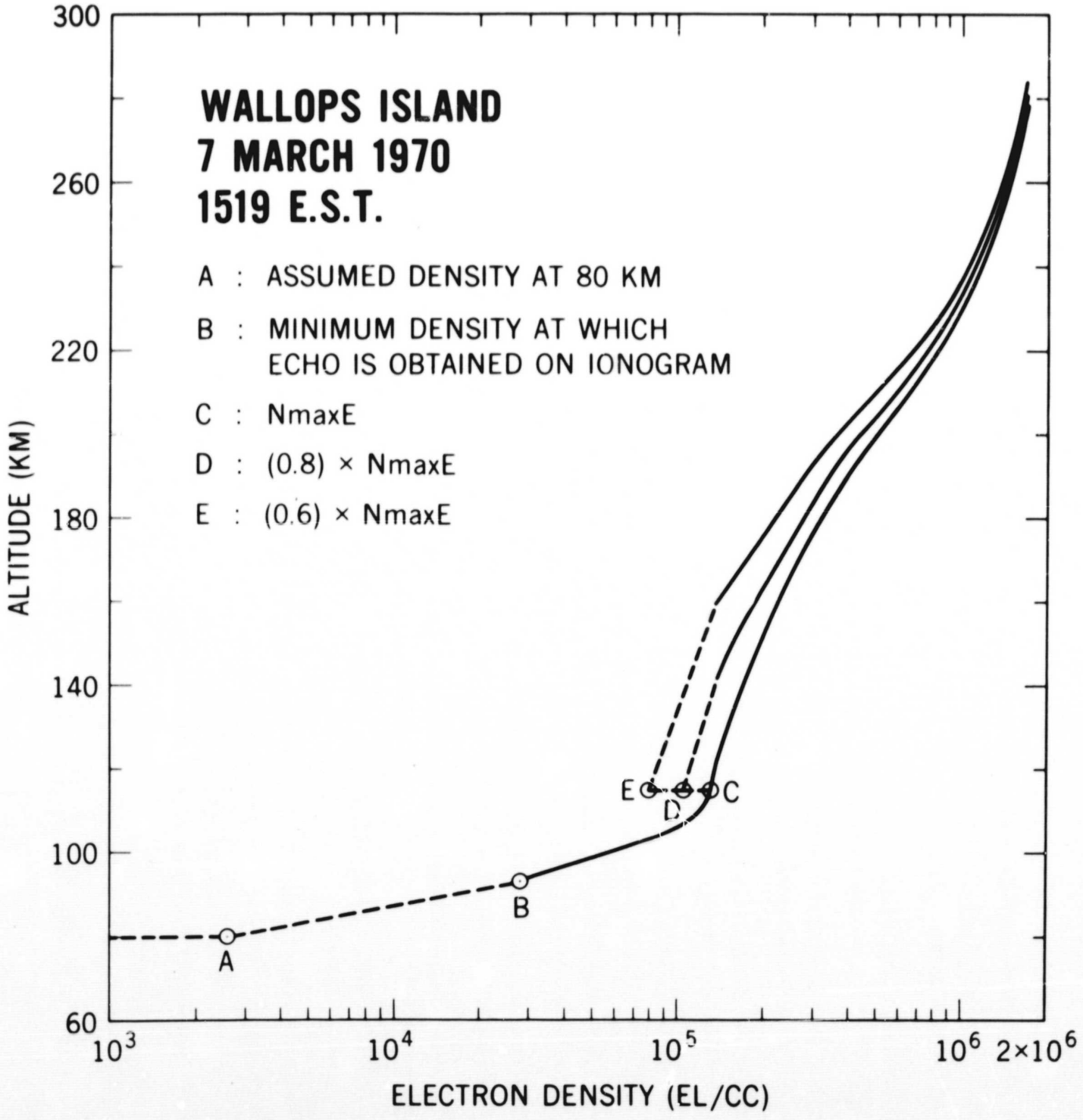
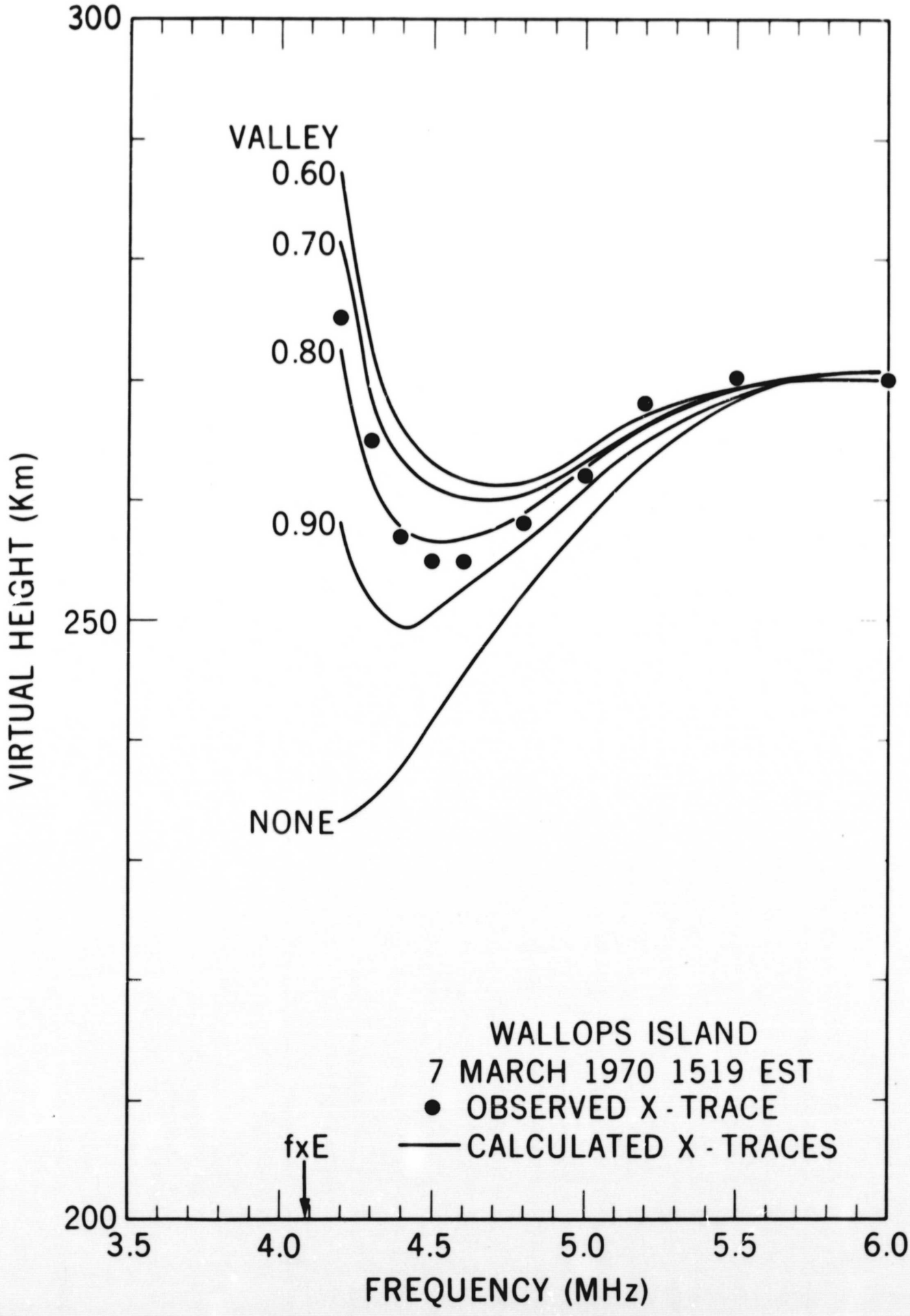

WALLOPS ISLAND
7 MARCH 1970
1519 E.S.T.
A : ASSUMED DENSITY AT 80 KM
B : MINIMUM DENSITY AT WHICH ECHO IS OBTAINED ON IONOGRAM
C : NmaxE
D : (0.8) × NmaxE
E : (0.6) × NmaxE
ALTITUDE (KM)
300
260
220
180
140
100
60
10^3
10^4
10^5
10^6
2×10^6
ELECTRON DENSITY (EL/CC)
A
B
C
D
E

Fig. 8


300
250
200
VIRTUAL HEIGHT (Km)
VALLEY
0.60
0.70
0.80
0.90
NONE
WALLOPS ISLAND
7 MARCH 1970 1519 EST
OBSERVED X-TRACE
CALCULATED X-TRACES
fxE
3.5
4.0
4.5
5.0
5.5
6.0
FREQUENCY (MHz)

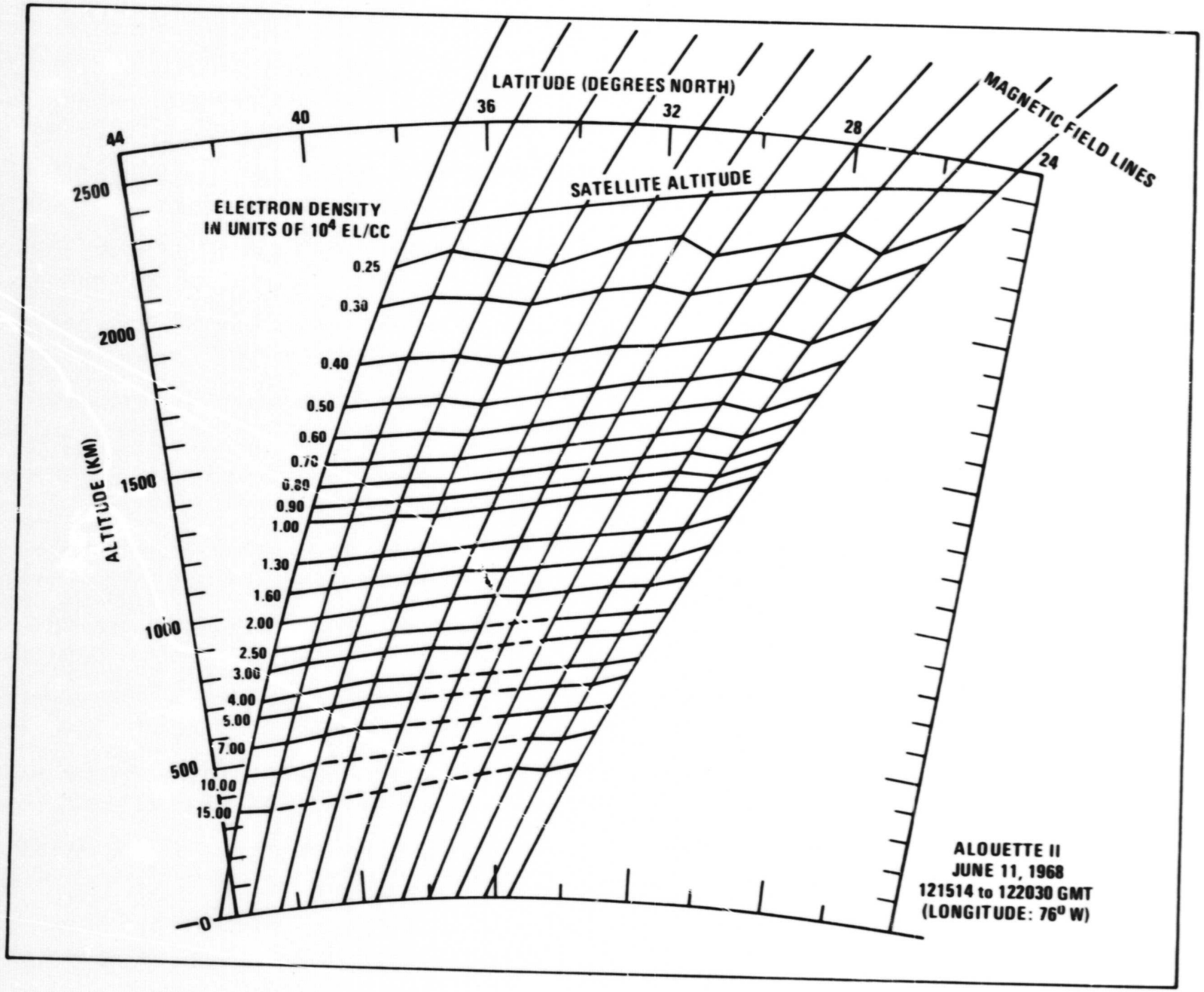

LATITUDE (DEGREES NORTH)
44
40
36
32
28
24
MAGNETIC FIELD LINES
SATELLITE ALTITUDE
ELECTRON DENSITY
IN UNITS OF 10^4 EL/CC
0.25
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
1.30
1.60
2.00
2.50
3.00
4.00
5.00
7.00
10.00
15.00
ALTITUDE (KM)
2500
2000
1500
1000
500
0
ALOUETTE II
JUNE 11, 1968
121514 to 122030 GMT
(LONGITUDE: 76° W)

Fig. 9

Fig. 10

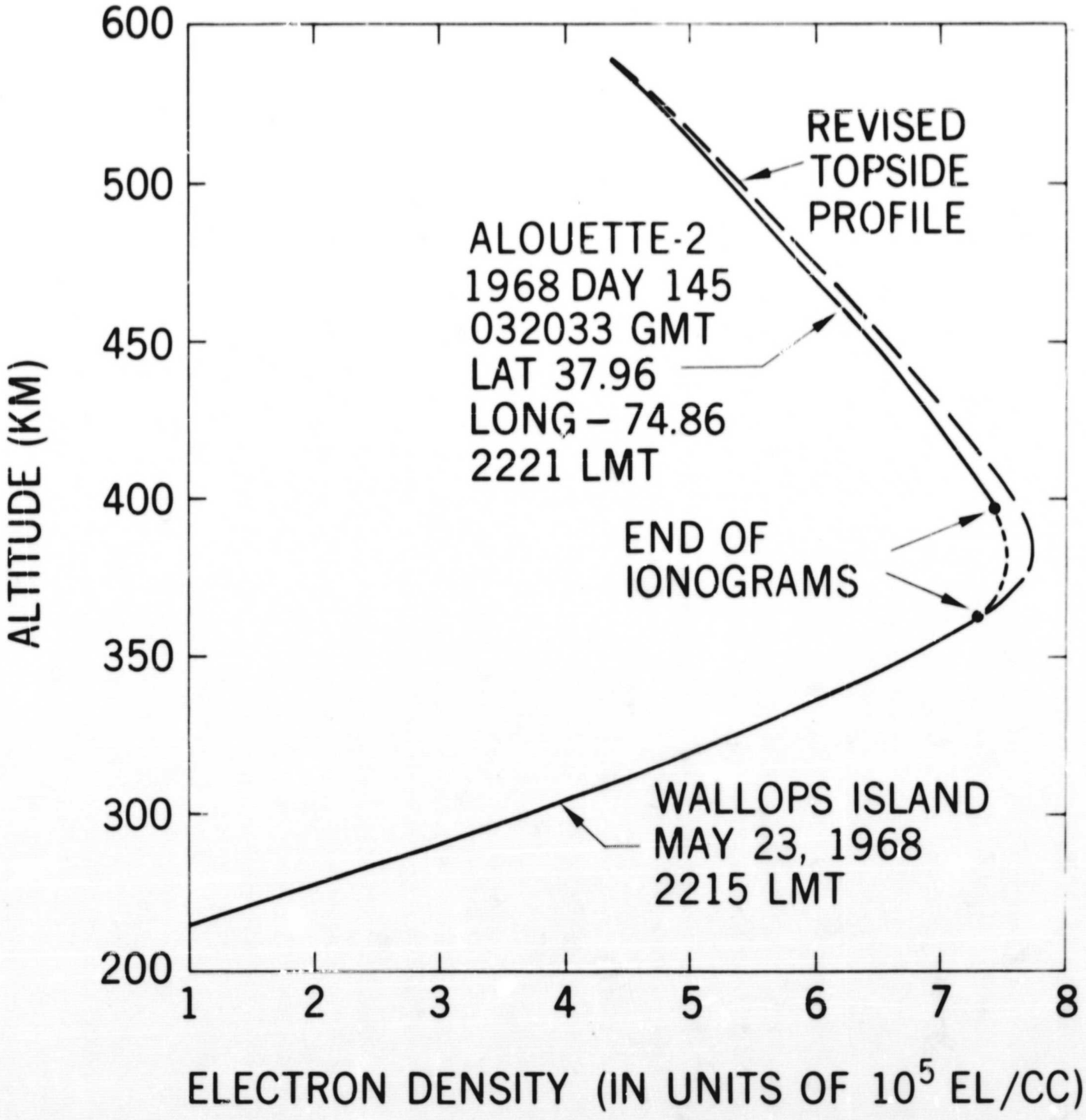
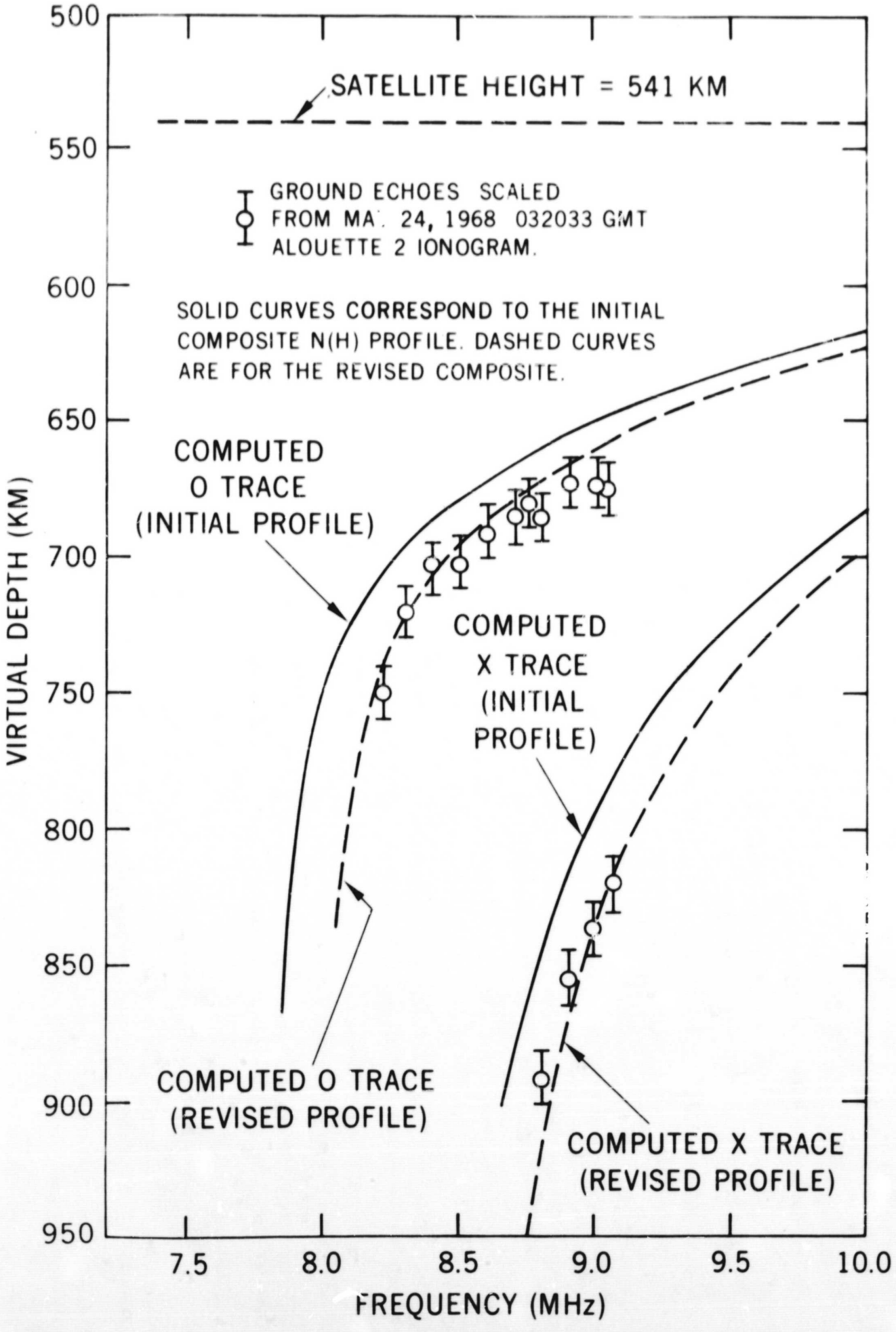

600
500
450
400
350
300
200
1
2
3
4
5
6
7
8
ALTITUDE (KM)
ELECTRON DENSITY (IN UNITS OF 10^5 EL/CC)
REVISED
TOPSIDE
PROFILE
ALOUETTE-2
1968 DAY 145
032033 GMT
LAT 37.96
LONG – 74.86
2221 LMT
END OF
IONOGRAMS
WALLOPS ISLAND
MAY 23, 1968
2215 LMT

Fig. 11


SATELLITE HEIGHT = 541 KM
GROUND ECHOES SCALED
FROM MAY 24, 1968 032033 GMT
ALOUETTE 2 IONOGRAM.
SOLID CURVES CORRESPOND TO THE INITIAL
COMPOSITE N(H) PROFILE. DASHED CURVES
ARE FOR THE REVISED COMPOSITE.
COMPUTED
O TRACE
(INITIAL PROFILE)
COMPUTED
X TRACE
(INITIAL
PROFILE)
COMPUTED O TRACE
(REVISED PROFILE)
COMPUTED X TRACE
(REVISED PROFILE)
VIRTUAL DEPTH (KM)
500
550
600
650
700
750
800
850
900
950
7.5
8.0
8.5
9.0
9.5
10.0
FREQUENCY (MHz)