# VIII. GENERAL

# "EXAMPLES OF CERTAIN DATA REDUCTION AND MAPPING PROCEDURES UTILIZING TIROS III FIVE-CHANNEL RADIOMETER DATA"*

LEWIS J. ALLISON AND GUENTER WARNECKE†

Selected examples of current data reduction and mapping procedures are discussed for the purpose of demonstrating their utility as well as their limitations.

## INTRODUCTION

In order to analyze and accurately interpret the TIROS radiation data, certain data reduction and mapping processes are described which should be of interest to the researcher and meteorologist alike.

A brief description of the 5-channel scanning radiometer is included for the benefit of the reader who may be unfamiliar with the TIROS radiation experiment.

## 1. THE FIVE CHANNEL SCANNING RADIOMETER

The TIROS III scanning radiometer consists of a five-channel instrument. The five channels are sensitive in the following spectral regions:[1]

| Channel | Spectral Region | |
|---|---|---|
| 1 | 6.0 to 6.5$\mu$ | (radiation within the $\nu_2$ fundamental water vapor band) |
| 2 | 8 to 12$\mu$ | (radiation emitted in the atmospheric "window") |
| 3 | 0.2 to 6$\mu$ | (total reflected solar radiation (albedo)) |
| 4 | 8 to 30$\mu$ | (total infrared radiation of the earth-atmosphere system) |
| 5 | 0.55 to 0.75$\mu$ | (reflected solar radiation within the visible part of the spectrum) |

The 5-channel radiometer employs a chopper that causes each sensor to view alternately and at a rapid rate, in two directions 180 degrees apart. At least one direction always views outer space which serves as a zero radiation reference. The response from each channel is proportional to the difference in the irradiation of the sensor bolometer from the two directions. When both optics view outer space, the resultant signal is designated as the "space viewed level". The bi-directional axes of the channels are parallel to one another and inclined to the spin axis by 45 degrees (Figure 1).[2]

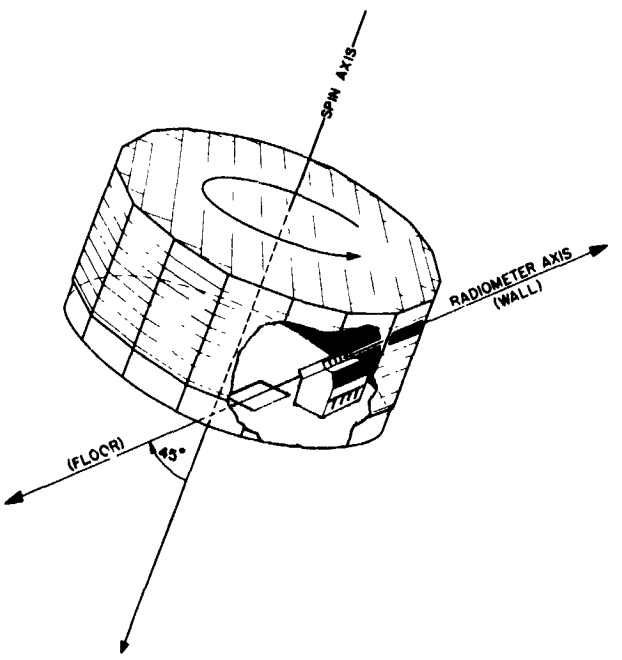


FIGURE 1.—Position of the Tiros Radiometer within the satellite.

*Published as *Goddard Space Flight Center Document X-651-64-132*, April 1964.

†Goddard Space Flight Center; on leave from the Institut für Meteorologie und Geophysik der Freien Universitat Berlin, Berlin, Germany as National Academy of Sciences – National Research Council Research Associate with the National Aeronautics and Space Administration.

## 2. REMARKS ON THE DATA REDUCTION AND MAPPING PROCEDURES

### 2.1 Equivalent Blackbody Temperature

The calibration of the infrared channels is in terms of equivalent temperatures, $T_{BB}$, of a blackbody target filling the 5° field of view, whereas the calibration of the solar channels is in terms of that portion of radiant emittance from a target filling the field of view to which the channel responds through its filters and other optical elements.[2]

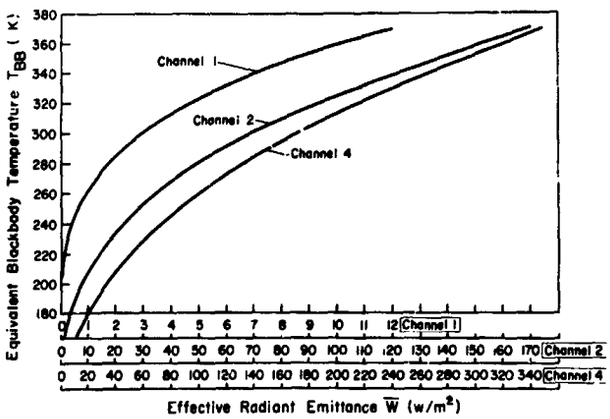


FIGURE 2.—The effective radiant emittance of the thermal channels of the Tiros III radiometer versus equivalent blackbody temperature.

The family of curves of equivalent blackbody temperatures vs effective radiant emittance for TIROS III (Figure 2), were obtained by integrating the Planck function over the spectral response curves for Channels 1, 2 and 4.[1,3] The estimated relative accuracy of equivalent blackbody temperature for Channels 1, 2 and 4 is ±2°K, while the estimated absolute accuracy is ±4° to ±8°K, up to orbits 118, 875 and 130 respectively, after which degradation of the instrumental response becomes appreciable.[1,4]

A degradation correction nomogram for Channel 2 measurements (Figure 3), based upon degradation models inferred from the relevant data obtained by the respective radiometer has been included.[4] The nomogram was developed from a computer program which determined the time rate of change of the measured total emitted radiant power between latitudes 55°N–55°S. The results showed that Channel 2 degraded continuously following the TIROS III launch. The degradation for Channel 2 was determined to be symmetrical, i.e., independent of the side (wall or floor aperture) of the satellite. The nomogram (Figure 3) may be used with some degree of confidence up through orbit 875. Beyond this subjectively determined limiting orbit, the correction nomogram becomes correspondingly uncertain. A more detailed discussion of the factors effecting the response of the scanning radiometer may be found in Reference 5.

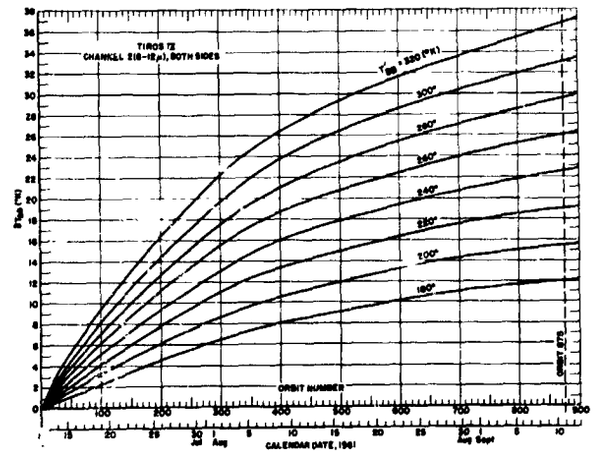


FIGURE 3.—The empirically-developed correction nomogram for instrumental degradation of Channel 2, Tiros III.

### 2.2 The Radiation Data Processing Procedure

The medium resolution radiometer data that are recorded at the read-out station are demultiplexed, demodulated and fed to an analog-to-digital converter at the Goddard Space Flight Center.[2] This converter produces a magnetic "Radiation Data Tape" made up of 36 bit words suitable for an IBM 7094 computer. Analog records may be produced on a recording oscillograph for special hand analysis. The computer produces a Final Meteorological Radiation Tape (FMRT) when combined with the inputs from the Radiation Data Tape, Orbital Tape and Calibration. The most convenient method of presentation of these FMRT data is to make grid point maps from an IBM 7094 computer. Several problems arising in the composition and averaging of the original radiation data are discussed below.

### 2.3 Radiation Data Resolution

The data resolution of computer produced radiation maps, printed on a mercator map base is directly affected by the following:

#### *2.3.1 Nadir Angle Variation*

Radiation data contained on the FMR Tapes are sampled approximately 8 times per second. Thus, at the nominal 10 rpm spin rate, approximately 48 samples are taken during each satellite rotation. Data sampled during the earth-viewed portion of the scan having the smallest nadir angle will be the most densely located, with progressively greater separation occurring at the higher nadir angles. Additionally, the size of the scan spot varies considerably with the nadir angle and with the height of the satellite. For example, at an 800 km satellite height (Figure 4) the elongated scan spot resolved by the radiometer at a 56.6° scan nadir angle, would enclose approximately 10 times the scan spot area at a scan-nadir angle of zero.[6] Thus the higher the radiometer nadir angle, the lower the data resolution.

Figure 5 shows that the nominal 5 degree field of view of the radiometer pertains to the 50 percent of peak-response level, and that there is some residual response to about ±4 degrees from the optical axis.[7] The percent power scan spots of Figure 4 are directly derived from this response curve.

#### *2.3.2 Grid Print Map Scale*

The mesh size of the multi-resolution mercator mapping program is 0.5 inches. Thus, the number of degrees of longitude per mesh interval is in effect the selection of the map scale. A resolution of 0.625°, 1.25° and 2.50° of longitude per mesh interval corresponds to a map scale of approximately 1:5 million, 1:10 million and 1:20 million, respectively, near the equator. On a mercator projection the map scale, the higher the

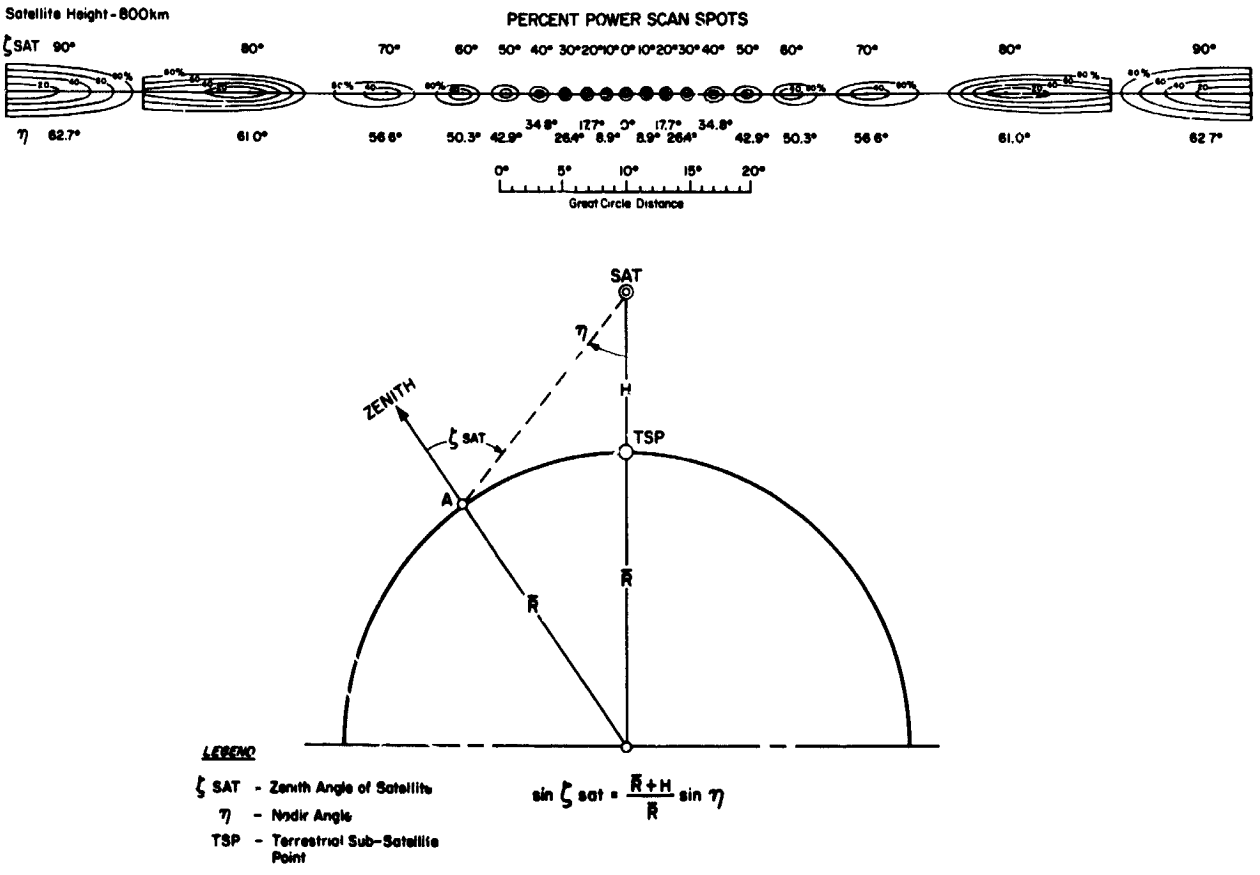


FIGURE 4.—Satellite zenith and nadir angle relationship. a.—Scan spot size as a function of zenith and nadir angle of the sensor and the power distribution within the scan spot. b.—Zenith angle and nadir angle relationship.

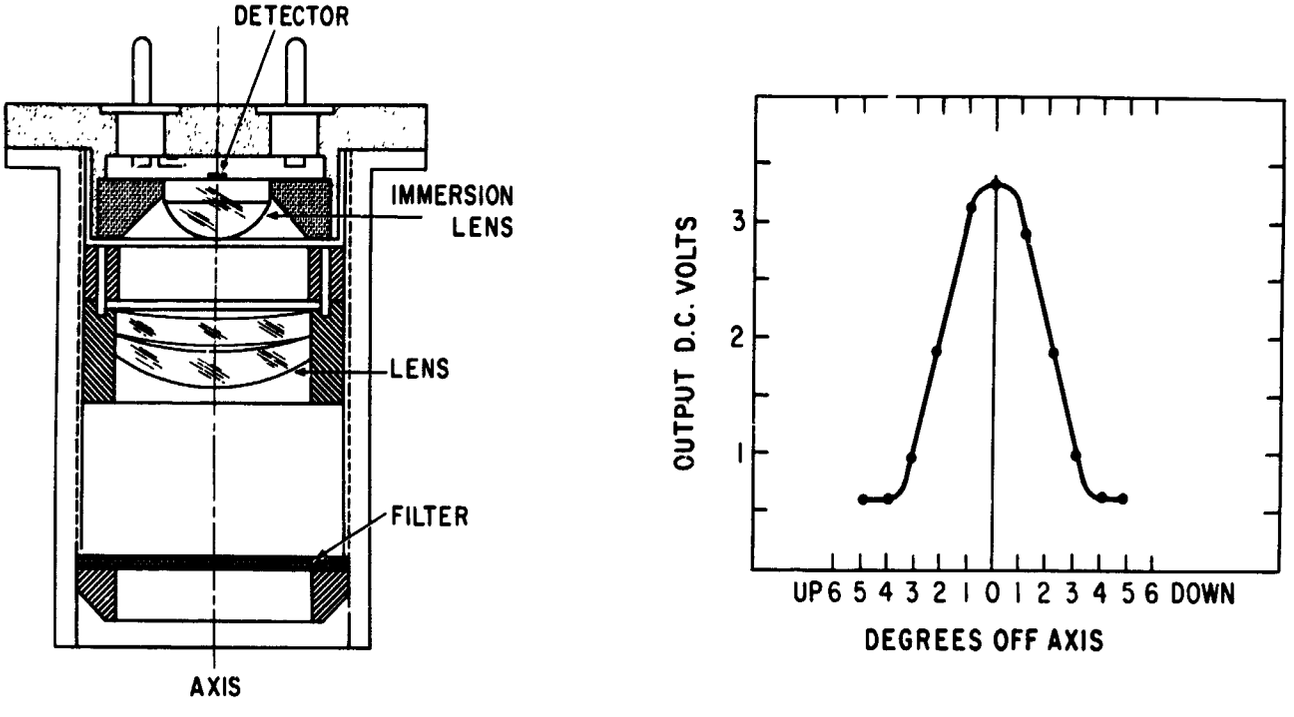


FIGURE 5.—Cross-section through a detector unit of the Tiros III radiometer and response curve within the field of view

resolution, the lower the number of measurements averaged with the ½ inch area around each grid point.

In order to illustrate the effect of map scale on data resolution, several TIROS III television pictures and radiation charts are shown in the following. The well-defined cloud system at 10°N, 40°W on July 16, 1961, which was identified by S. Fritz[8] as being the cloud mass from which Hurricane Anna, later developed in July, 1961, was selected as the example.

Figure 6a shows the "window" radiation analysis based upon hand-plotted data from a computer listing, which contains the geographical location, the nadir and azimuth angles, and the simultaneous readings of all five radiation channels for every single scan spot. The location of each scan spot is indicated by a cross in Figure 6a. This is the most detailed information available from the FMR Tape.

Figure 6b is a composite of the "window" radiation field shown in the previous figure and the gridde TIROS photograph of the same area. The latitude-longitude grid of this picture was obtained from a computer program.

The agreement between the picture information and the radiation field in the "window" region is surprisingly good, even at the fringes of the

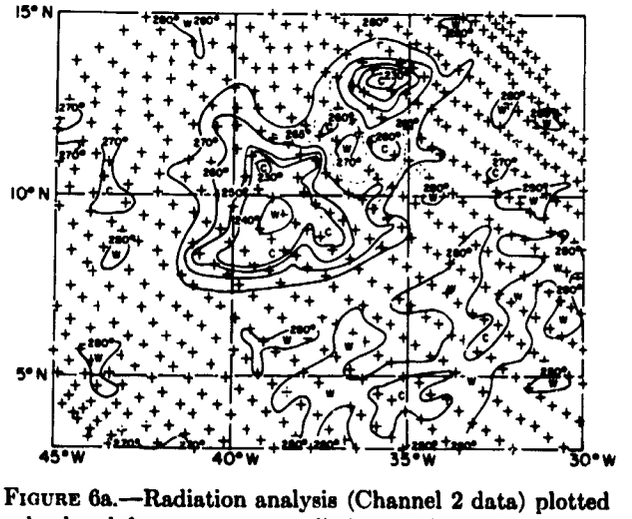


FIGURE 6a.—Radiation analysis (Channel 2 data) plotted by hand from a computer listing of the early stage of Hurricane Anna, which was used for the composite in Figure 6b.

photograph. Considering the range of accuracy of satellite attitude determination and the computer gridding programs, it is interesting to note that the radiation data fit the television pictures within one degree of latitude and longitude. An equivalent blackbody temperature difference of more than 50°K was indicated between the cloudless area and the highest cloud tops.

A similar radiation analysis (Figure 6c) was later performed by L. F. Whitney, Jr. of the

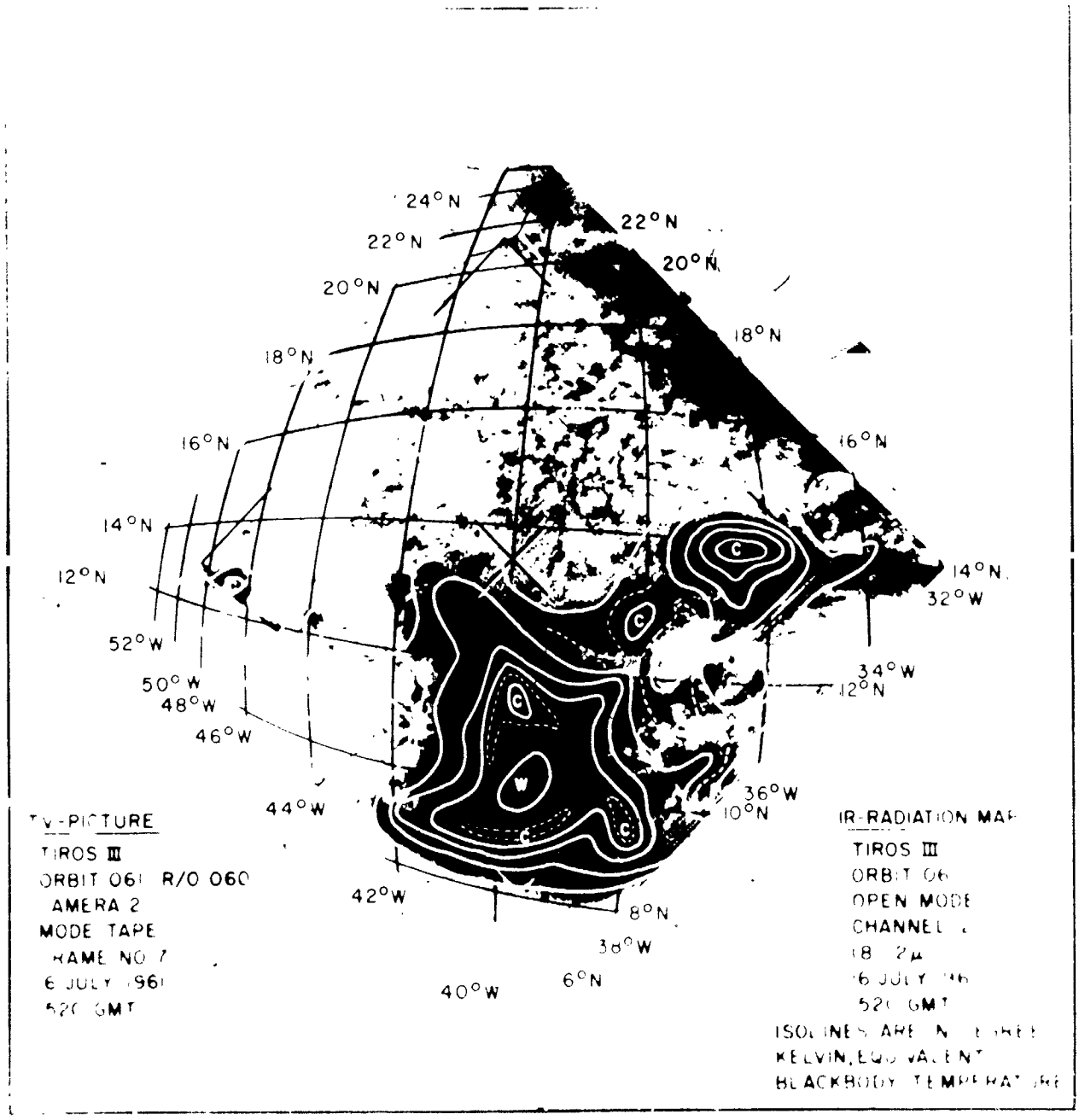


FIGURE 6b.—Composite analysis of Channel 2 radiation data and a gridded television picture of the early stage of Hurricane Anna.

National Weather Satellite Center, U.S. Weather Bureau, using the highly accurate Fujita method.[6] This method employs analog records produced on a recording oscillograph for special hand analysis rather than the automatically processed FMR Tapes. It should be considered, however, that this picture represents another frame of the same series, one minute later in time. From Figures 6b and 6c one can judge, that, at least in this case, the evaluation of the FMRT data does not result in any significant loss of details when compared with the results of the Fujita technique.

Figure 7 shows the radiation analysis of the same storm area based on three different types of computer produced grid maps at three different grid spacings, viz., 0.625°, 1.25°, and 2.50° longitude per mesh interval, corresponding to approximate map scale of 1:5 million, 1:10 million and

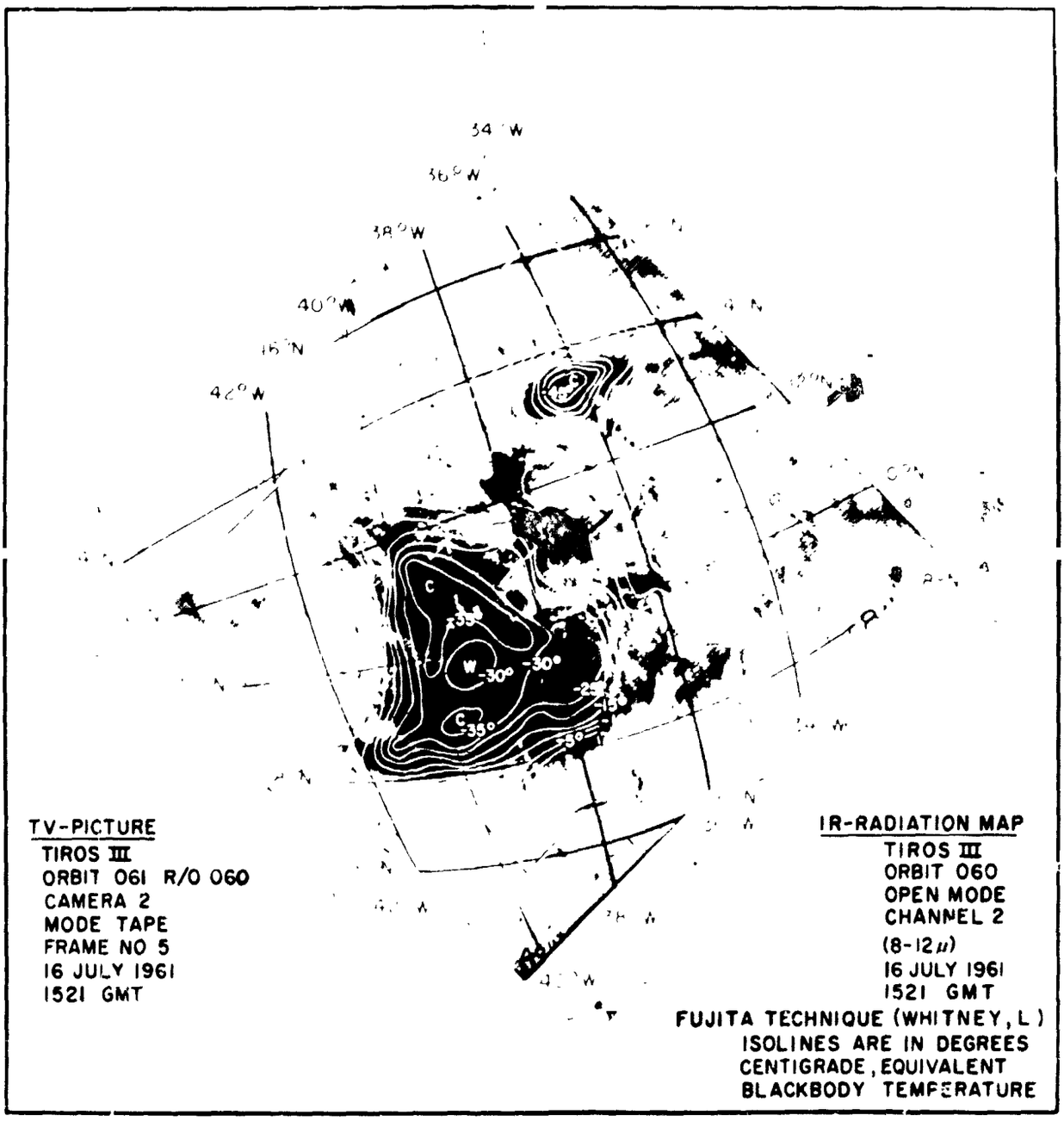


FIGURE 6c.—Composite analysis of Channel 2 radiation data and a gridded television picture of the early stage of Hurricane Anna, using the Fujita Technique by L. F. Whitney, Jr., U.S. Weather Bureau.

1:20 million respectively. The data points are again indicated in each figure. From the left to the right it becomes clear that the appropriate selection of the map scale for each radiation study is of great importance, as detailed features in the isoline configuration and extreme values of equivalent blackbody temperatures become "washed out" by the averaging processes involved at smaller map scales. Thus, for the analysis of more detailed synoptic features, a grid spacing of 0.625° of longitude per mesh interval is the most useful one, but the accuracy in data location is considerably less than by use of hand-plotted maps of the FMRT computer listings and of maps produced by the more precise Fujita method.

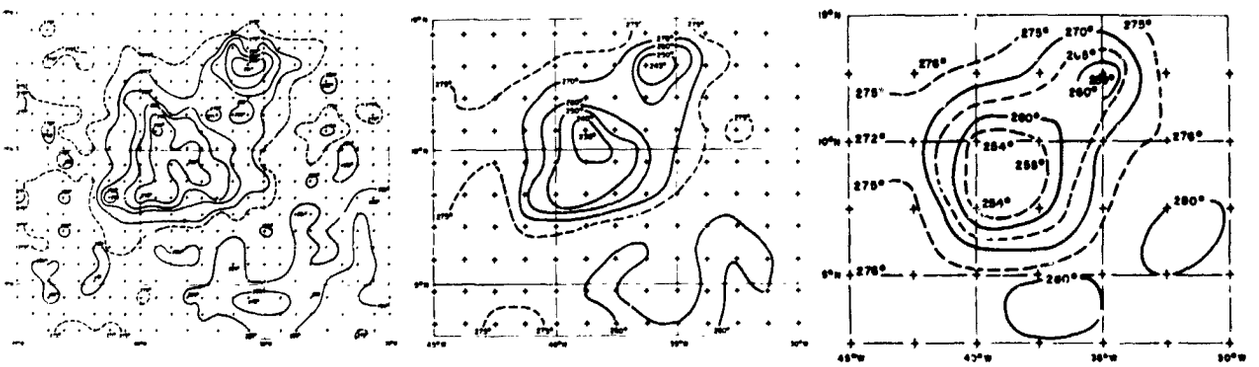

FIGURE 7.—Analysis of Channel 2 radiation data of the early stage of Hurricane Anna, using different resolutions in computer-produced grid print maps (Open mode data only).
a.—0.625° longitude/mesh interval
b.—1.250° longitude/mesh interval
c.—2.50° longitude/mesh interval

### 2.4 Radiometer Scanning Modes

Figure 1 showed the basic relationships between the satellite's spin axis and the radiometer axis.

As the satellite rotates on its axis, the radiometer scan pattern is defined by the intersection of a 45° half-angle cone and the earth. This pattern ranges from nearly a circle (closed mode) when the satellite nadir angle is less than17°, to two hyperbola-like branches in the "wall" and "floor" directions, when the spin vector is approximately perpendicular to the orbital radius vector.[1]

Therefore the three patterns of earth scanning modes, (Figure 8) can be distinguished as follows.[9]

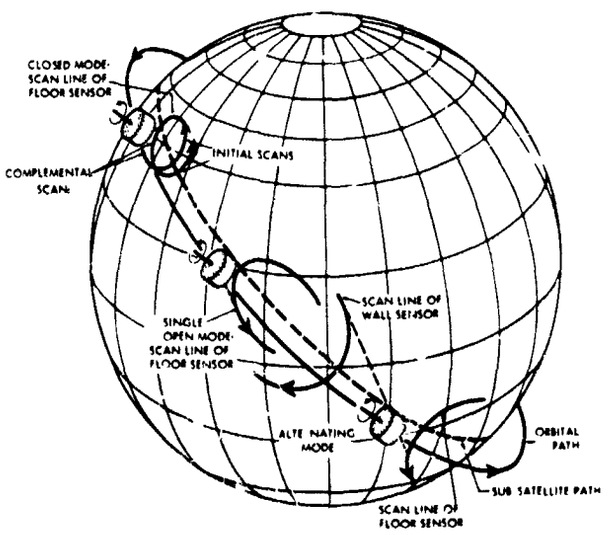


FIGURE 8.—Different types of scan modes of the Tiros radiometer generally occurring in one orbit.

1. Single Open Mode: Some scan spots of a spin cycle are space viewed and the remainder are earth viewed through the wall (side) sensor only or through the floor (baseplate) sensor only.

2. Closed Mode: All the scan spots throughout a number of spin cycles of the satellite are earth-viewed, either through the wall or the floor sensor.

3. Alternating Open Mode: The scan spots of a spin cycle are alternately viewed through the wall sensor and the floor sensor.

The IBM 7094 computer can easily be programmed to distinguish between different modes and to use or omit certain scanning modes. For example, in Figure 9 for orbits 56 and 58, those parts with single open mode data are grey shaded, and it is obvious that large and important parts of the data coverage are lost if closed mode data are omitted. Some examples of actual scan lines and scan spots are indicated on orbit 56. Additionally in this figure, the computer grid print map of orbit 57 is introduced, showing the full data coverage along the orbital path. However, there is a significant difference in accuracy of the geographical location between open and closed mode data, which reduces the applicability of closed mode data unless a very detailed investigation of possible short-term aberrations of the pseudo-spin rate value used in the computer reduction of the closed mode data is performed for every individual case. In the following

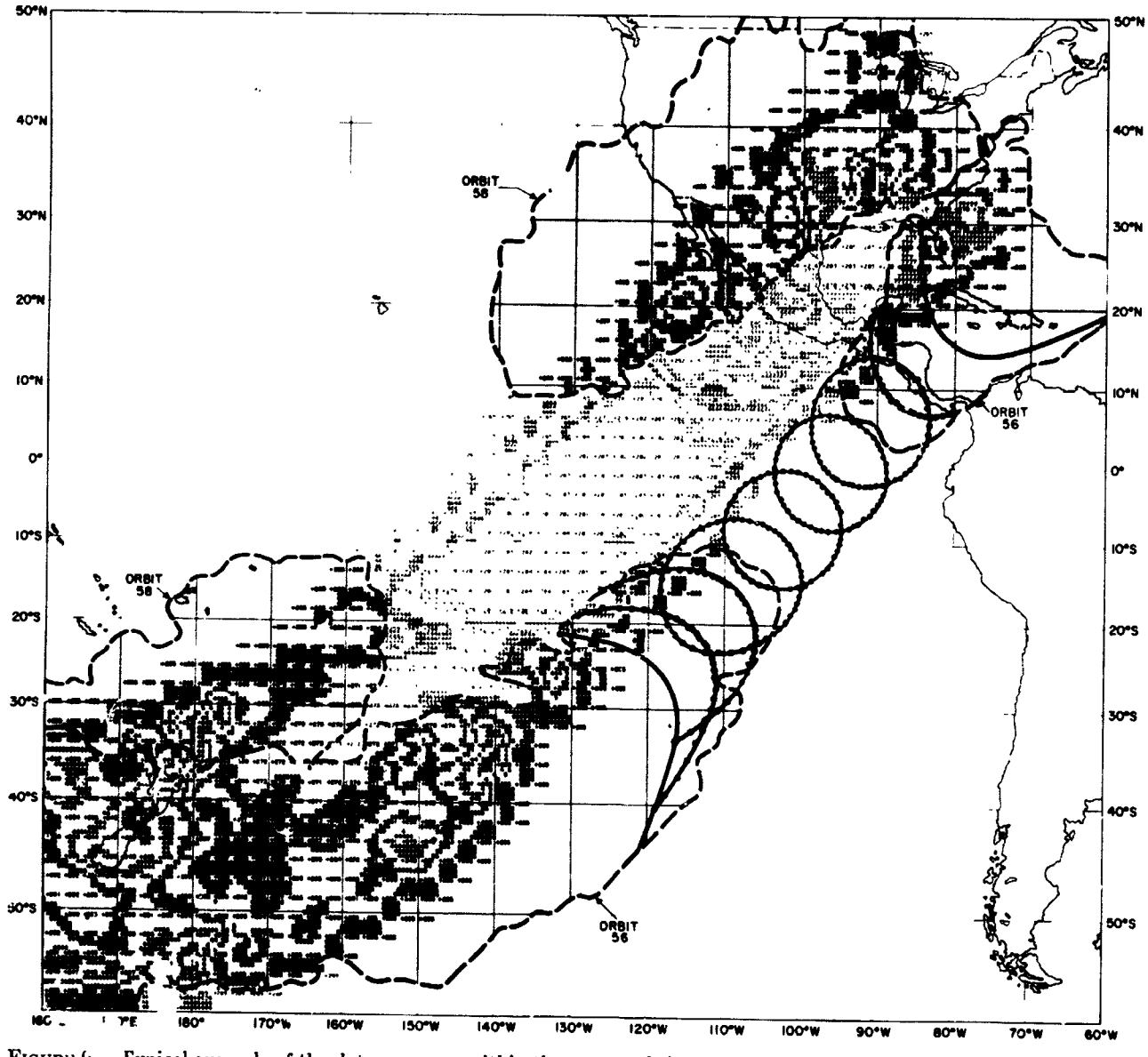


FIGURE 9.--Typical example of the data coverage within the open and closed scanning modes for three orbits of Tiros III. Some typical scan lines and the distribution of computer-selected scan spots along these lines are added for demonstration of some characteristic features of scan line geometry.

sections these practical difficulties will be discussed more in detail.

An example of typical closed mode data mislocation is given in Figure 10. This figure includes 10 minutes of closed mode data and shows the radiation "cloud" pattern of the early stage of Hurricane Anna at the same location within minutes of the same time as shown in Figure 7, which included open mode data only. This longer time period permitted the complemental scans of the closed mode to follow by several minutes the initial scans through the storm area (see Figure 8), resulting in an apparent shift of data leading to the figure. The floor aperture of the satellite was viewing the storm at this time, and a 6 scan spot shift in a counterclockwise direction was confirmed by diagnostic methods. The computer program averaged the displaced scan spot data with the preceding accurately positioned open mode scan spots and presented and erroneously elongated cloud structure.

A separate example of analog presentations of open and closed mode patterns is shown in Figures 11 and 12. Line A identifies the measurement

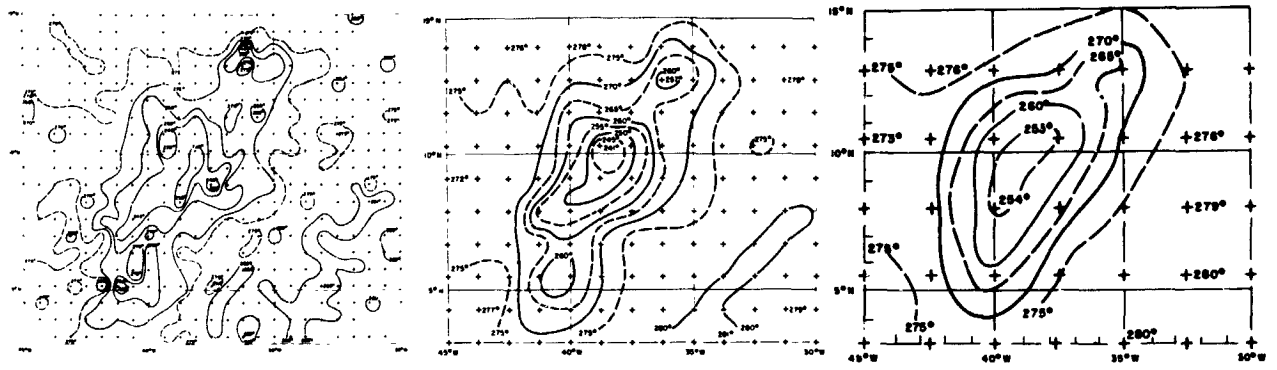

FIGURE 10.—Analysis of Channel 2 radiation data of the early stage of Hurricane Anna, using different resolutions in computer-produced grid print maps. (Open and closed mode data).
a.—0.625° longitude/mesh interval
b.—1.250° longitude/mesh interval
c.—2.50° longitude/mesh interval

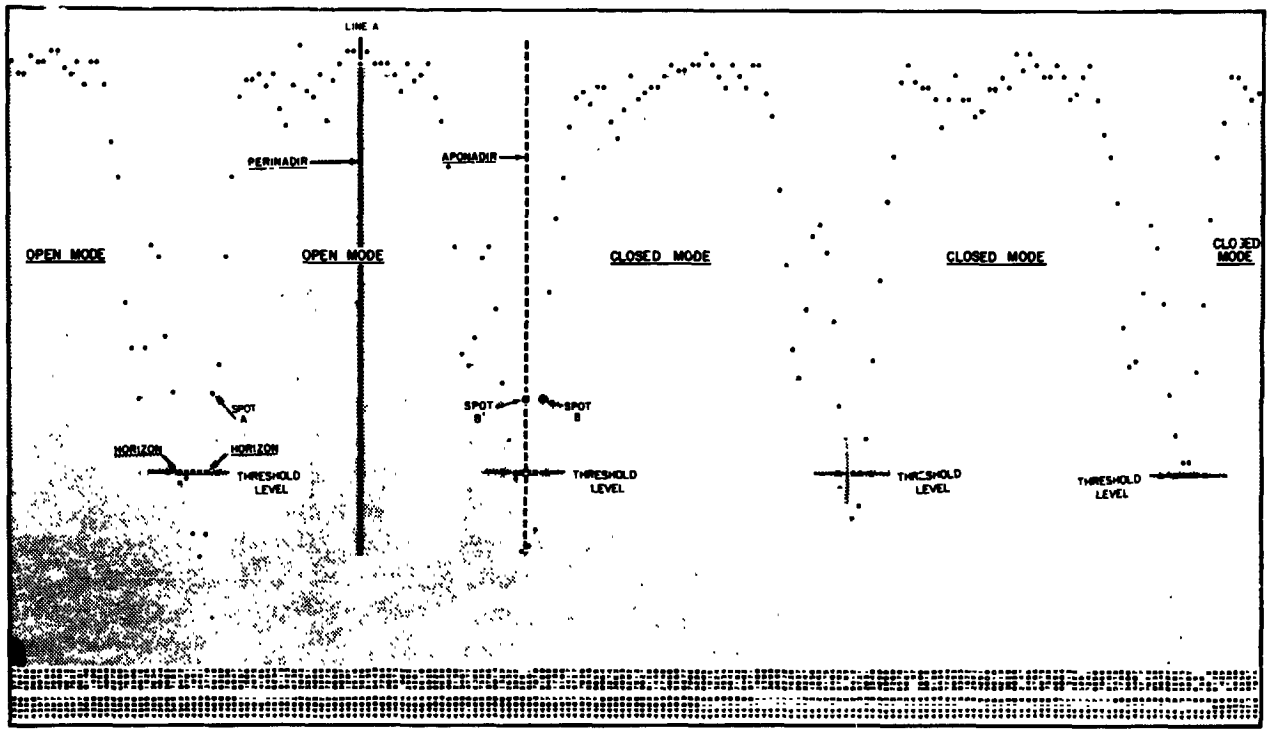


FIGURE 11.—Digitized form of the analog signal from Channel 2 during the transition from the single open to closed scanning modes. The numbers along the bottom indicate the minutes and seconds from the beginning of the data. The ordinate is the radiometer output on a relative scale.

obtained when viewing the perinadir in an open mode swath. This line is centered between the horizons of the swath and all the scan spots are positioned symmetrically about this line by the computer program, in accordance with the spin rate. Normally there can be little or no error in positioning these scan spots since each swath is processed individually and the horizons for each are determined quite accurately. For the closed mode swaths, mislocation errors do occur simply because there are no horizons detected by the satellite and the location of the data must be determined solely by the spin rate.

The spin rate used in the data reduction program is measured by means of counting cycles of a standard clock at the read-out station between

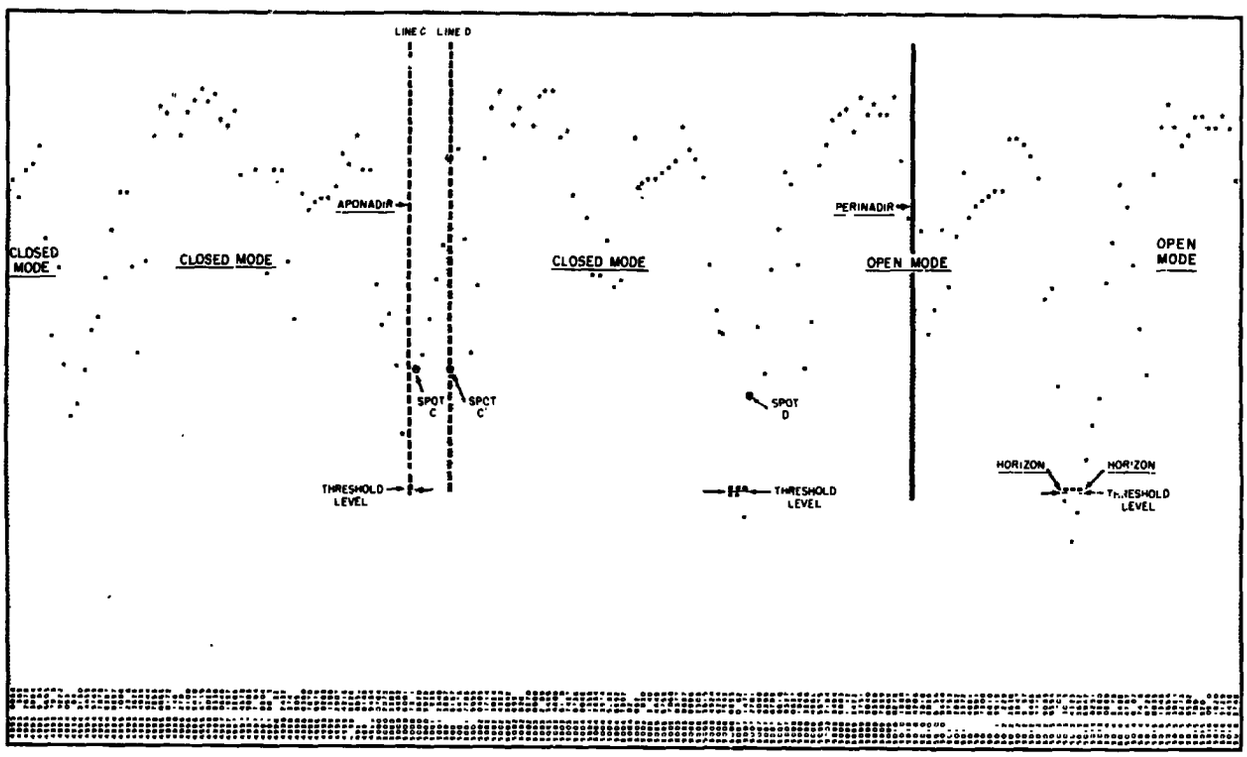


FIGURE 12.—Digitized form of the analog signal from Channel 2 during the transition from closed to single open scanning modes. The numbers along the bottom indicate the minutes and seconds from the beginning of the data. The ordinate is the radiometer output on a relative scale.

sun pulses and is assumed to be constant throughout the closed mode. For a spin rate determination error of 0.01°/sec (or 1 part in 6000), the spin angle error at the end of 10 minutes would be 6°, or about equal to one scan spot.[9]

The computer program is such that the first scan spot (Spot B) after the last open mode swath is positioned on the aponadir (Spot B'), a shift backward of 2½ scan points or about 15° of the spin angle.

All succeeding scan spots in the swath are shifted proportionally which results in a shift of the radiation data by 2½ scan spots (approx. 200 km). Figure 13 illustrates this mislocation in relation to geographic position, giving, however, examples for points close to the centers of the swaths. The measurement at point $H$ of the first closed mode swath, which should be contiguous to that of point $G$ within the last open mode swath, will be mislocated by the computer and appears in the wrong position $H'$. Comparing the last closed with the first open mode swath, point $I$ should be contiguous to the measurement at point $J$; in this instance however, it was mislocated by the computer and shifted into the wrong location $I'$.

The 1½ to 2½ scan spot shift that occurs at the beginning of the closed mode is always counter to the rotation of the swath. If the spin rate used in the computer program is less than the actual rate, the mislocation of successive scan spots will be counter to the scan rotation. If the assumed spin rate is greater than the actual, the mislocation will be in the direction of the scan rotation.

The spin rate used for the 7 orbits in this study were all less than the actual rate, so that by the end of the closed mode, the mislocation averaged six data points or about 36° of the spin angle counter to the direction of scan rotation. Therefore, all scan spots were shifted backwards so that by the end, they were 6 spots too far in back of the true position. Thus, for example, spot C′ was erroneously located on the aponadir, (Figure 12) instead of on its correct position which should have been 6 spots ahead of the aponadir.

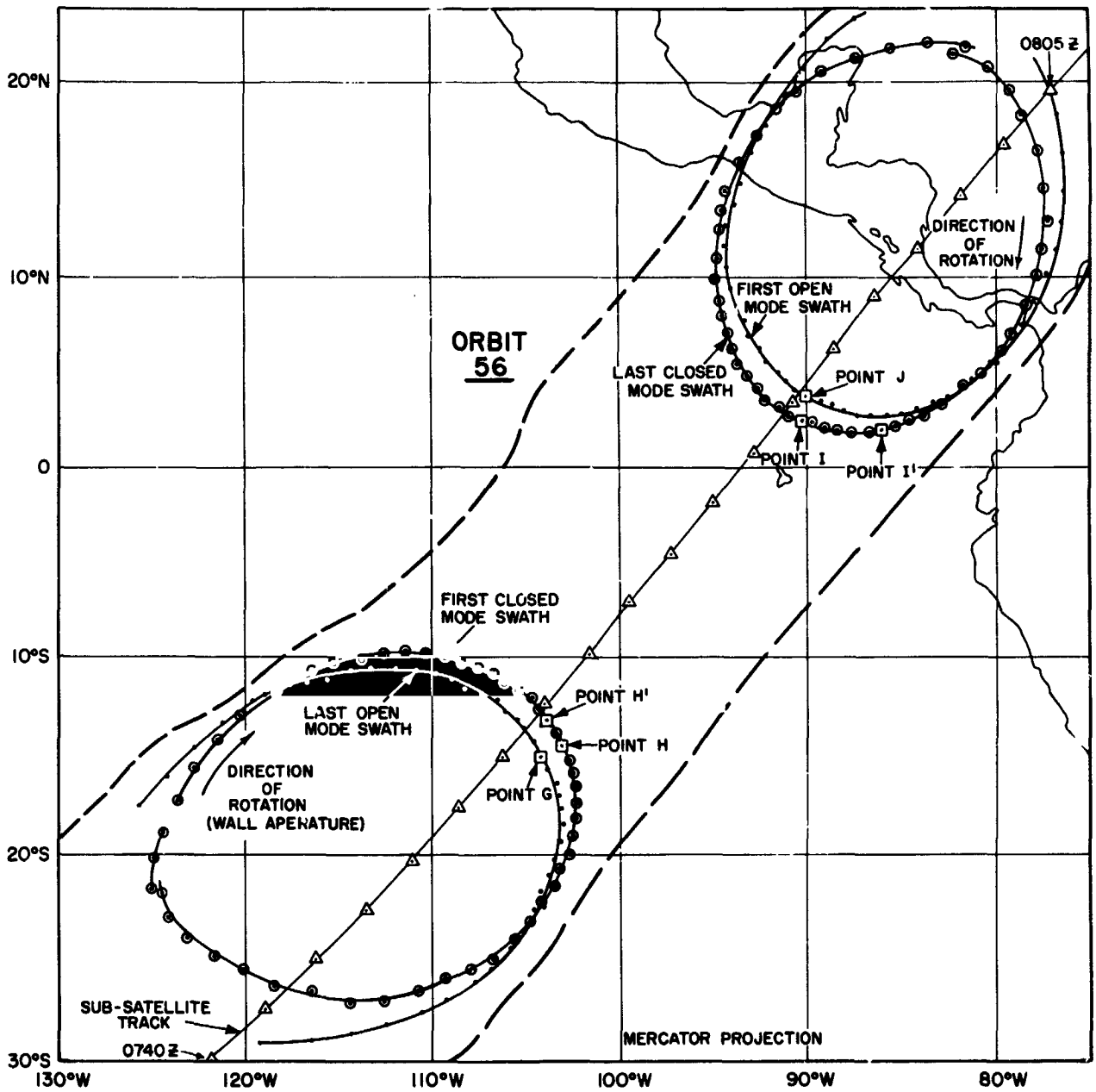


FIGURE 13.—An example of the scan spot shift at the entrance and exit from the closed scanning mode due to computer program techniques and inaccuracy in spin rate determination.

In order to improve the accuracy of location of the closed mode data, steps have been initiated to calculate the psuedo-spin rate values which should be individually applied to each closed mode sequence in the automatic reduction of the data.

### 2.5 Measurements Falling Below the Space-earth Discriminant

For the reasonable interpretation of the recorded radiometer signals it is necessary to distinguish very carefully between earth-scanned data and space-viewed or space-contaminated data. Therefore, the computer logic was set to make the distinction by means of an arbitrarily defined threshold value below which the data should be defined as "space viewed". Theoretically, this space-earth discriminant should be identical with the space viewed level and be located far from all atmospheric data. Practically, however, it is necessary to define the threshold value above the noise level of the radiometer system. Normally, there is no interference with

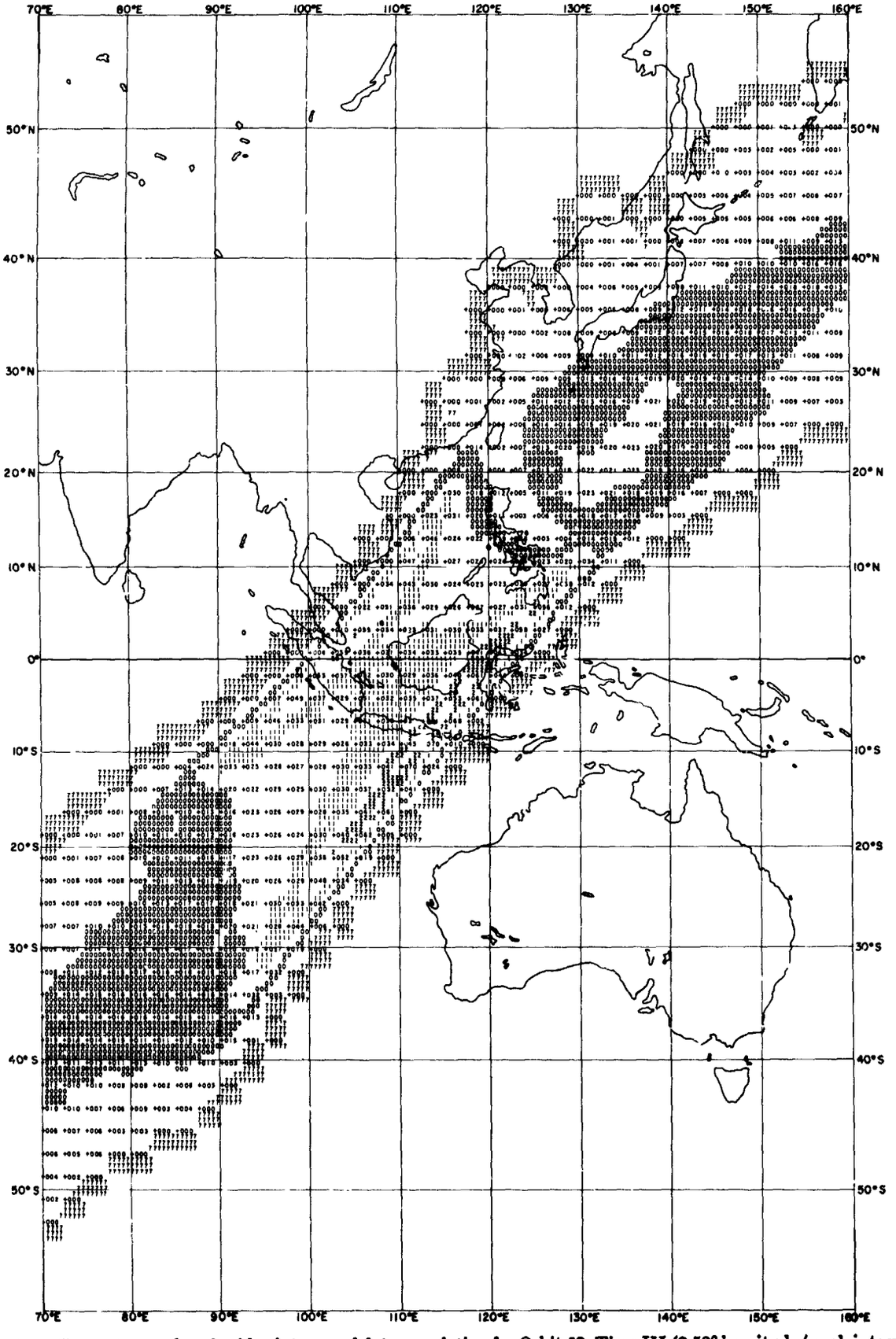


FIGURE 14.—Computer-produced grid print map of data population for Orbit 62, Tiros III (2.50° longitude/mesh interval).

real atmospheric measurements. But in the case of very low radiation intensities the actual equivalent blackbody temperatures may be lower than the threshold value. This occurs mainly over areas of very high opaque cloud systems such as tropical storms.

If a measurement appears below the space-earth discriminant, according to the TIROS III Radiation Data Users' Manual[1], it will be flagged by a minus sign on the FMR Tape and in the computer listings. Three or more consecutive values below the threshold will be assumed to define the end of a swath by the computer logic. Also, all data making up the terminal swath of the closed mode are flagged with minus signs for diagnostic purposes. This means that over large high-cloud areas or when an unusually high noise level occurs just in the transition between closed and open scanning modes, the computer logic may not be able to distinguish properly between real earth-scanned and space-viewed data, resulting in a large number of measurements being flagged with minus signs.

Generally, the computer will be ordered to reject all data flagged by a minus sign in order to avoid the use of questionable data. The com-

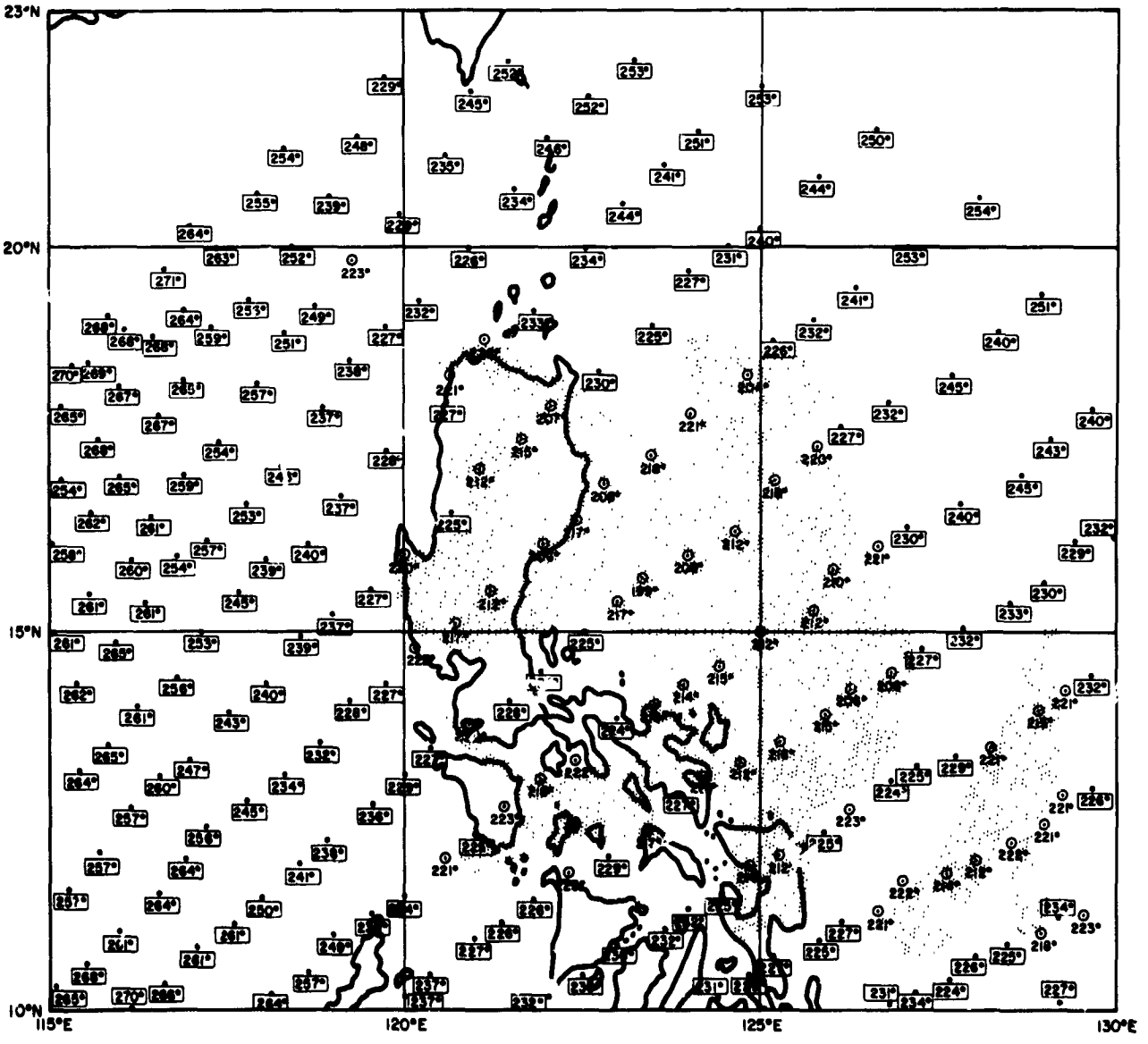


FIGURE 15.—Location of scan spots over a tropical disturbance close to the Philippines, based upon computer listings. (Encircled data in the shaded area were rejected by computer logic in producing the grid print maps).

puter has, however, an option to accept these data for grid print maps. Therefore, a careful check of the FMR Tape data is necessary under such circumstances to permit the analyst in advance to select the areas where the flagged data are meaningful for mapping.

An example of such an occurrence is given in Figure 14. By inspecting the population count, which shows how many data points are contributing to each grid point in the computer produced grid print map, for orbit 62 an irregularity is noted over the Philippine Islands. Approximately 20 data points were rejected by the computer in every grid square in this area. Figure 15 shows all the single data points taken from the computer listings. All data accepted by the computer mapping program are marked by a rectangle, while the rejected "flagged" data points are encircled. The final (Figure 16) analysis of this radiation field shows that just the most interesting low radiation temperature data (very high reaching dense cloud system) over Tropical Storm Flossie had been rejected. Thus, the effect of the artificially defined space-earth discriminate on the composition of radiation data over certain meteorologically significant areas

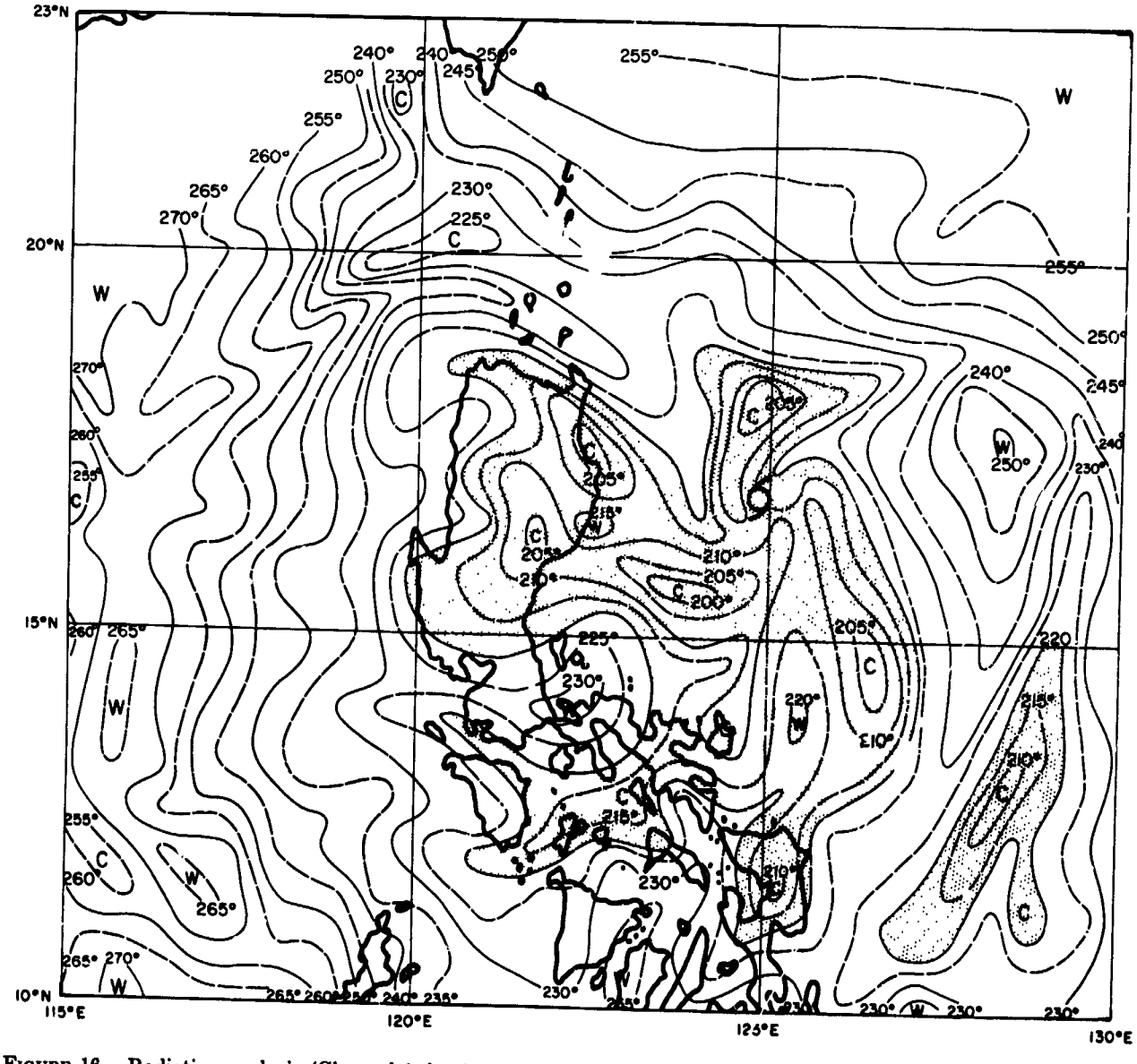


FIGURE 16.—Radiation analysis (Channel 2 data) over a tropical disturbance close to the Philippines, based upon computer listings. (Data within the grey shaded area were rejected by computer logic in producing the grid print maps.)

should always be accounted for in detailed studies. In certain cases, perhaps, the space-earth discriminant could be set somewhat lower, but this depends on such conditions, as signal-to-noise ratio, radiometer housing temperatures, etc. which are discussed in Reference 10.

## 3. CONCLUSIONS

This paper has attempted to review briefly the satellite radiation experiment and to discuss the current TIROS radiation data reduction processes. With improvement and modification of the computer processing procedures, it will become possible to reduce the location (closed mode) uncertainties of the data and also to accept low but real measurements so that these data could be used in operational weather analyses.

## REFERENCES

1. TIROS III Radiation Data Users' Manual, Goddard Space Flight Center, NASA, August 1962.
2. BANDEEN, W. R., "Data Processing From Meteorological Satellites," NASA SP-16, Nat'l. Aero. and Space Admin., Dec. 1962.
3. NORDBERG, W., W. R. BANDEEN, B. J. CONRATH, V. KUNDE, and I. PERSANO, "Preliminary Results of Radiation Measurements from the TIROS III Meteorological Satellite," Jrl. Atmosp. Sci., 19, 20–30, 1962.
4. BANDEEN, W. R., R. E. SAMUELSON, I. P. STRANGE, "TIROS III Radiation Data Users' Manual Supplement, Correction Models for Instrumental Response Degradation," NASA, Goddard Space Flight Center, Dec. 1, 1963.
5. BANDEEN, W. R., V. KUNDE, W. NORDBERG, and H. P. THOMPSON, "TIROS III Meteorological Satellite Radiation Observations of a Tropical Hurricane," Tellus—to be published (1964).
6. FUJITA, T., "Outline of a Theory and Examples for Precise Analysis of Satellite Radiation Data," Research Paper 15, University of Chicago, Feb. 1963.
7. ASTHEIMER, R. W., R. DE WAARD, and E. A., JACKSON, "Infrared Radiometric Instruments on TIROS II", Jrl. of Optical Society, 51, 1386–93, Dec. 1961.
8. FRITZ, S., "Satellite Pictures and the Origin of Hurricane Anna," Monthly Weather Review, Vol. 90, pp. 507–513, Dec. 1962.
9. BANDEEN, W. R., "TIROS II Radiation Data Users' Manual Supplement," NASA Goddard Space Flight Center, Greenbelt, Md., May 15, 1962.
10. TIROS IV Radiation Data Catalog and Users' Manual, NASA/Goddard Space Flight Center, Greenbelt, Md., 15 December 1963.

## ACKNOWLEDGEMENT

The authors are indebted to Mr. W. R. Bandeen, Mr. George W. Nicholas, Mr. Harold P. Thompson, of Goddard Space Flight Center and Mr. L. F. Whitney, Jr., U. S. Weather Bureau for their contributions to the presented analysis work.

# SYMMETRIC EULER ANGLE DECOMPOSITION OF THE TWO-ELECTRON FIXED-NUCLEUS PROBLEM*

A. K. BHATIA† AND A. TEMKIN

## 1. INTRODUCTION

The decomposition of the Laplacian operator,

$$\nabla^2=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\left(\frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\sin\vartheta\frac{\partial}{\partial\vartheta}+\frac{1}{\sin^2\vartheta}\frac{\partial^2}{\partial\varphi^2}\right),$$

where the coefficient of the second term is proportional to the square of the angular momentum operator, is the basic relation between kinetic energy and angular momentum in the quantum mechanics of the one body problem (or the relative motion of two particles). When acting on a wave function which is an eigenfunction of total angular momentum $\ell$, the Laplacian simplifies to

$$\nabla^2=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)-\frac{\ell(\ell+1)}{r^2},$$

in which form it is clear that the effect of this decomposition is to reduce the Schödinger equation with a spherically symmetric potential from a three dimensional partial differential equation to a one dimensional (ordinary) differential equation. As such this relation is of fundamental mathematical importance and is familiar to everyone who utilizes quantum mechanics at all.

The analogous procedure when more than one particle is involved, in particular two identical particles in an external force field, although known, is not as well known, nor is it as well developed. When the external field is that of a fixed nucleus the wave function is expanded in eigenfunctions of the total angular momentum of the two particles multiplied by functions of the three remaining independent variables. The total angular momentum eigenfunctions are functions of the three Euler angles only. These angles are not unique, but in some way they must describe the orientation of the instantaneous plane formed by the two particles and the center of coordinates (nucleus) in space. The remaining three coordinates then describe the positions of the particles in this plane, and the functions of these variables are the generalized radial functions. Hylleraas' original papers[1] in effect contained the reduced or radial equations for total S states in terms of the residual coordinates $r_1$, $r_2$, $r_{12}$. In this case, the total orbital angular momentum is a constant function, and hence the reduction of a six-dimensional to a three-dimensional partial differential equation is independent of how one defines the Euler angles.

The standard treatment of the general problem is due to Breit.[2,3] He used the Euler angles that Hylleraas [1] originally introduced: namely the two spherical angles of one of the particles in the space fixed coordinate system and a second azimuthal angle between the $r_1$–$z$ plane and the $r_2$–$r_1$ plane. Breit's remaining coordinates were chosen as $r_1$, $r_2$, and $\theta_{12}$, the latter being the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$. If, however, one wants to describe two electron atoms or ions in the approximation that the nuclei are fixed, then one has an additional requirement of which there is no analogue in the one-body problem. And that is the Pauli principal: the requirement that the spatial function be either symmetric or antisymmetric under the exchange of the particle coordinates. It is clear that the Hylleraas-Breit choice of Euler angles (which we hereinafter refer to as the Hylleraas-Breit angles), being quite unsymmetrical with respect to the two particles, is not optimum in this respect.[4] In fact the construction the linear combinations of angular momentum functions with the appropriate exchange properties is a very difficult task which

*Published as *Goddard Space Flight Center Document X-640-64-68*, March 1964.

†National Academy of Sciences—National Research Council Resident Research Associate.

depends not only on the Euler angles but on $\theta_{12}$ as well. It is not surprising, therefore, that Breit's original work[2] was limited to $P$-states, and work thereafter has always been limited to specific angular momentum states.[5]

However, a treatment by Holmberg[6] using a symmetrical choice of Euler angles (which we shall call Holmberg's angles) has in the interim been carried out. With these angles the description of exhange as well as parity (which latter property is also simply describably with th *II-B* angles) is simple (although these properties are only alluded to in Holmberg's paper). One of the purposes of the present paper is to examine these properties and relate them more clearly to the construction of the total wave function and further clarify other aspects of Holmberg's important paper. Consequently, we shall be enabled to derive the general radial equations for arbitrary angular momentum ($\ell$) for the case of two identical particles in the field of a fixed nucleus. Holmberg's treatment applies to three particles of the same mass.

It is clear that in treating three particles of equal mass Holmberg had in mind the three nucleon problem whereas we are interested in two-electron atoms, ions, and diatomic molecules. The application of this formalism to two electron atoms and ions is clear, and the decomposition amounts to a rigorous reduction of the Schrödinger equation. It should only be remarked here that the scattering of electrons from one electron atoms and ions is also a special class of these problems. We have therefore worked out the connection between the boundary conditions for electron-atom scattering and Holmberg's angles (Section VIII).

Inasmuch as Holmberg's paper refers to the three nucleon problem, reference should also be made to the papers of Derrick and Blatt.[7] These papers deal much more realistically with the three nucleon problem in that full account is taken of an internucleon potential which is considerably more complicated than a central potential. As regards the actual choice of coordinates, Derrick and Blatt define axes along the moments of inertia of the three-body system. As such th[e] will depend on the lengths of the interparticle distances and therefore are quite different fr[o]m Holmberg's angles (Section II).

## II. HOLMBERG'S ANGLES

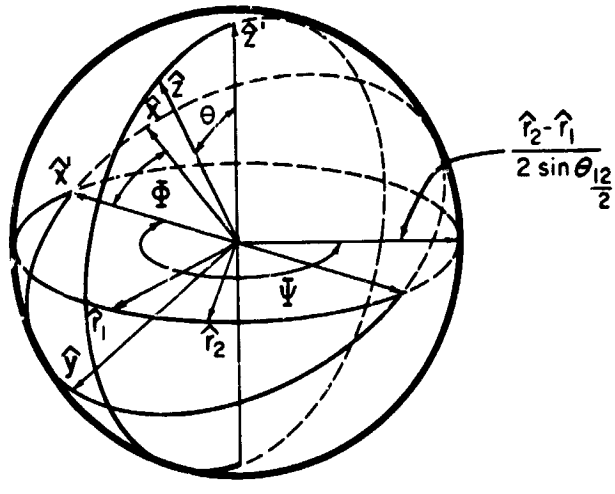


FIGURE 1.—Perspective drawing of (Holmberg's) Euler angles and the unit vectors of the problem.

Figure 1 contains a perspective drawing of Homberg's angles which define the particle plane with respect to the space fixed $x$, $y$, and $z$ axes. The rotated axes $x'$, $y'$, $z'$ are then defined by

$$\hat{z}'=\frac{\hat{r}_1\times\hat{r}_2}{|r_1\times r_2|} \tag{1}$$

$$\hat{x}'=\frac{\hat{z}'\times\hat{z}'}{|z\times z'|} \tag{2}$$

$$\hat{y}'=\hat{z}'\times\hat{x} \tag{3}$$

The Holmberg Euler angles are then

$$\theta\equiv\text{angle between } \hat{z} \text{ and } \hat{z}' \tag{4}$$

$$\Phi\equiv\text{angle between } \hat{x}' \text{ and } \hat{x} \tag{5}$$

$$\Psi\equiv\text{angle between } \hat{x}' \text{ and } (\hat{r}_2-\hat{r}_1) \tag{6}$$

The ranges and planes of these angles are:

| | |
|---|---|
| $0\leq\theta\leq\pi$ | in $z$–$z'$ plane |
| $0\leq\Phi\leq2\pi$ | in $x$–$y$ plane |
| $0\leq\Psi\leq2\pi$ | in $x'$–$y'$ plane |

As is usual a cap on a vector is used to represent a unit vector in the given direction. In particular $\hat{i}$, $\hat{j}$, $\hat{k}$ are the three unit vectors along the (space fixed), $x$, $y$, and $z$ axes respectively, and thus are synonymous with $\hat{x}$, $\hat{y}$, and $\hat{z}$. Similarly $\hat{i}'$, $\hat{j}'$, $\hat{k}'$ and $\hat{x}'$, $\hat{y}'$, $\hat{z}'$ are identical.

It is clear from the figure that $\hat{x}$, being in the $x$–$y$ plane has components:

$$\hat{x}' = \hat{i}\cos\Phi + \hat{j}\sin\Phi \tag{7}$$

Since $\hat{x}'$ is perpendicular to the $z$–$z'$ plane, it is perpendicular to every line in that plane going through the origin. This includes specifically the line of intersection of the $z$–$z'$ plane with the $x$–$y$ plane. However, the azimuthal angle of that intersecting line is the azimuth of $\hat{z}'$ itself, and since $\hat{x}'$ has azimuth $\Phi$, $\hat{z}'$ has azimuth $\frac{3\pi}{2}+\Phi$ (cf. Figure 1). The polar angle of $\hat{z}'$ is clearly $\theta$, therefore we have the important relation:

$$\hat{z}' = \hat{i}\sin\theta\sin\Phi - \hat{j}\sin\theta\cos\Phi + \hat{k}\cos\theta \tag{8}$$

The relations between Holmberg's angles and the spherical angles of the individual particles are obtained by substituting Eqs. (7) and (8) into the *lhs* of Eqs. (1) and (2) and using the ordinary decomposition of $\mathbf{r}_1$ and $\mathbf{r}_2$ in the *rhs*:

$$\hat{r}_1 = \hat{i}\sin\vartheta_1\cos\varphi_1 + \hat{j}\sin\vartheta_1\sin\varphi_1 + \hat{k}\cos\vartheta_1$$

$$\hat{r}_2 = \hat{i}\sin\vartheta_2\cos\varphi_2 + \hat{j}\sin\vartheta_2\sin\varphi_2 + \hat{k}\cos\vartheta_2$$

One obtains

$$\sin\theta_{12}\cos\theta = \sin\theta_1\sin\vartheta_2\sin(\varphi_1-\varphi_1) \tag{9}$$

$$\sin\theta_{12}\sin\theta\sin\Phi = \sin\vartheta_1\sin\varphi_1\cos\theta_2 - \cos\vartheta_1\sin\vartheta_2\sin\varphi_2 \tag{10}$$

$$\sin\theta_{12}\sin\theta\cos\Phi = \sin\vartheta_1\cos\varphi_1\cos\vartheta_2 - \cos\vartheta_1\sin\vartheta_2\cos\varphi_2 \tag{11}$$

$$2\sin(\tfrac{1}{2}\theta_{12})\cos\psi = \sin\varphi_2\cos(\varphi_2-\Phi) - \sin\vartheta_1\cos(\varphi_1-\Phi) \tag{12}$$

$$2\sin(\tfrac{1}{2}\theta_{12})\sin\psi = (\cos\vartheta_2-\cos\vartheta_1)\sin\theta - \sin\theta_1\sin(\varphi_1-\Phi) + \sin\vartheta_2\sin(\varphi_2-\Phi) \tag{13}$$

$$\cos\theta_{12} = \cos\vartheta_1\cos\vartheta_2 + \sin\vartheta_1\sin\vartheta_2\cos(\varphi_1-\varphi_2) \tag{14}$$

The latter relation is, of course, the well known expansion for the angle between two vectors.

It is also of interest to give the vectors $\hat{r}_1$ and $\hat{r}_2$ in the particle plane: (primed co-ordinate system):

$$\hat{r}_1 = \hat{i}'\sin(\psi-\tfrac{1}{2}\theta_{12}) - \hat{j}'\cos(\psi-\tfrac{1}{2}\theta_{12}) \tag{15}$$

$$\hat{r}_2 = \hat{i}'\sin(\psi+\tfrac{1}{2}\theta_{12}) - \hat{j}'\cos(\psi+\tfrac{1}{2}\theta_{12}) \tag{16}$$

$$\hat{z} = \hat{j}'\sin\theta + \hat{k}'\cos\theta \tag{17}$$

$$\hat{y} = \hat{i}'\sin\Phi + \hat{j}'\cos\theta\cos\Phi - \hat{k}'\sin\theta\cos\Phi \tag{18}$$

$$\hat{x} = \hat{i}'\cos\Phi - \hat{j}'\cos\theta\sin\Phi + \hat{k}'\sin\theta\cos\Phi \tag{19}$$

The following relations which are also very useful can now simply be derived by computing $(\hat{r}_1\cdot\hat{z})$, $(\hat{r}_2\cdot\hat{z})$ etc. in the primed system.

$$\cos\vartheta_1 = -\sin\theta\cos(\psi-\tfrac{1}{2}\theta_{12}) \tag{20a}$$

$$\cos\vartheta_2 = -\sin\theta\cos(\psi+\tfrac{1}{2}\theta_{12}) \tag{20b}$$

$$\sin\vartheta_1\cos\varphi_1 = \cos\Phi\sin(\psi-\tfrac{1}{2}\theta_{12}) + \cos\theta\sin\Phi\cos(\psi-\tfrac{1}{2}\theta_{12}) \tag{21a}$$

$$\sin\vartheta_2\cos\varphi_2 = \cos\Phi\sin(\psi+\tfrac{1}{2}\theta_{12}) + \cos\theta\sin\Phi\cos(\psi+\tfrac{1}{2}\theta_{12}) \tag{21b}$$

$$\sin\theta_1\sin\varphi_1 = \sin\Phi\sin(\psi-\tfrac{1}{2}\theta_{12}) - \cos\theta\cos\Phi\cos(\psi-\tfrac{1}{2}\theta_{12}) \tag{22a}$$

$$\sin\vartheta_2\sin\varphi_2 = \sin\Phi\sin(\psi+\tfrac{1}{2}\theta_{12}) - \cos\theta\cos\Phi\cos(\psi+\tfrac{1}{2}\theta_{12}) \tag{22b}$$

## III. PROPERTIES UNDER PARITY AND EXCHANGE

The operation of parity corresponds to the simultaneous inversion of both particles' coordinates: $\mathbf{r}_1 \to -\mathbf{r}_1$, $\mathbf{r}_2 \to -\mathbf{r}_2$. It can be seen from figure 1 that this places $\mathbf{r}_1$ and $\mathbf{r}_2$ facing the opposite direction, but the cross product and hence $\hat{z}'$ will not change as a result of this operation. Thus the $z$–$z'$ plane will not change and $\hat{x}'$ will not change. On the other hand $(\hat{r}_2-\hat{r}_1)$ goes into the negative of itself, so that $\psi$ gets increased by $\pi$. In other words under *parity*

$$\begin{aligned}\theta &\to \theta\\ \Phi &\to \Phi\\ \psi &\to \psi+\pi\end{aligned} \tag{23}$$

Exchange corresponds to the transformation $\mathbf{r}_1 \rightleftarrows \mathbf{r}_2$. From the analytical definitions $\hat{z}'$ and $\hat{x}'$, Eqs. (1) and (2), the new primed axes will go into the negative of themselves. Also $(\hat{r}_2-\hat{r}_1)$ goes into negative of itself. Clearly the inversion of the $z'$

axis corresponds to the transformation $\theta\to\pi-\theta$. Noting that $\Phi$ is the angle in the $x$–$y$ plane and measured as positive with respect to the $z$ axis, which is fixed, we see that $\Phi\to\pi+\Phi$. The simultaneous inversion of $\hat{\mathbf{x}}'$ and $(\mathbf{r}_2-\mathbf{r}_1)$ means that the modulus of the angle $\psi$ remains the same. However, since $\psi$ is an angle in the $x'$–$y'$ plane, it is measured as positive with respect to the $z'$ axis. Since the latter goes into the negative of itself, it becomes clear that $\psi\to2\pi+\psi$. Thus we have under *exchange*

$$\begin{aligned}\theta&\to\pi-\theta\\ \Phi&\to\Phi+\pi\\ \psi&\to2\pi+\psi\end{aligned}\tag{24}$$

The significance of these transformations relates to the transformation properties of the vector spherical harmonics under the same operation. These functions, which are the eigenfunctions of the angular momentum (next section), are the basic functions in terms of which the complete wave function is expanded. They can be written

$$D_l^{m,k}(\theta,\Phi,\psi)=\frac{\sqrt{2(2l+1)}}{4\pi}e^{i(m\Phi+k\psi)}d_l^{m,k}(\theta)\tag{25}$$

where the normalization has been so chosen that the function is identical with what is given in Section IV and the $d_l^{mk}(\theta)$ agree with those given by Wigner.[8] Only the dependence on $\theta$ is nontrivial:

$$d_l^{m,k}(\theta)=(-1)^{\frac{1}{2}(|k-m|+k-m)}\frac{4\pi}{\sqrt{2(2l+1)}}N_{lmk}\sin^{|k-m|}\frac{\theta}{2}\times$$
$$\times\cos^{|k+m|}\frac{\theta}{2}F\left(-l-\frac{\beta}{2}-1,\ l+\frac{\beta}{2},\ 1+|k-m|,\ \sin^2\frac{\theta}{2}\right),\tag{26}$$

where $\beta$ and $N_{lmk}$ are defined in Eqs. (45) and (46). $F(a,b;c;z)$ is the hypergeometric function in the notation of Magnus and Oberhettinger.[9] The important property of $d_l^{mk}(\theta)$, proved in Wigner's book,[8] is:

$$d_l^{m,k}(\pi-\theta)=(-1)^{l-m}d_l^{m,-k}(\theta)\tag{27}$$

The phase in (26) is such that this relation holds for all values of $m$ and $k$.

Letting $P$ and $E_{12}$ represent parity and exchange, we have from (23) and (24):

$$PD_l^{m,k}(\theta,\Phi,\psi)=D_l^{m,k}(\theta,\Phi,\pi+\psi)$$

$$E_{12}D_l^{m,k}(\theta,\Phi,\psi)=D_l^{m,k}(\pi+\theta,\pi+\Phi,2\pi-\psi)$$

which using (27) reduce to

$$PD_l^{m,k}(\theta,\Phi,\psi)=(-1)^kD_l^{m,k}(\theta,\Phi,\psi)\tag{28}$$

$$E_{12}D_l^{m,k}(\theta,\Phi,\psi)=(-1)^lD_l^{m,-k}(\theta,\Phi,\psi)\tag{29}$$

The simplicity of Eq. (29) is the essential feature which recommends Holmberg's angles to the description of the two electron problem.

## IV. ANGULAR MOMENTUM

The components of the total angular momentum are readily expressed in terms of the particles spherical angles. Thus, for example

$$\frac{-i}{\hbar}M_z=\sin\varphi_1\frac{\partial}{\partial\theta_1}+\cot\vartheta_1\cos\varphi_1\frac{\partial}{\partial\varphi_1}+\sin\varphi_2\frac{\partial}{\partial\vartheta_2}$$
$$+\cot\vartheta_2\cos\varphi_2\frac{\partial}{\partial\varphi_2}\tag{30}$$

The particles' angles $\vartheta_1$, $\varphi_1$, $\vartheta_2$, $\varphi_2$ via (9)–(14) are implicit functions of the four angles $\theta$, $\Phi$, $\psi$, $\theta_{12}$. Thus the problem of finding $M_z$ in these angles is a straightforward problem of partial differentiation. We can write

$$\frac{-i}{\hbar}M_z=A_\theta\frac{\partial}{\partial\theta}+A_\Phi\frac{\partial}{\partial\Phi}+A_\psi\frac{\partial}{\partial\psi}+A_{\theta12}\frac{\partial}{\partial\theta_{12}}$$

where

$$A_\chi=\sin\varphi_1\frac{\partial\chi}{\partial\vartheta_1}+\sin\varphi_2\frac{\partial\chi}{\partial\vartheta_2}+\cot\vartheta_1\cos\varphi_1\frac{\partial\chi}{\partial\varphi_1}$$
$$+\cot\vartheta_2\cos\varphi_2\frac{\partial\chi}{\partial\varphi_2}\tag{31}$$

and $\chi$ can be any one of the angles $\theta$, $\Phi$, $\psi$, or $\theta_{12}$.

Using then Eqs. (9)–(14), one finds that the following relations fall out quite easily:

$$A_{\theta 12}=0 \tag{32}$$

$$A_\theta=-\cos\Phi \tag{33}$$

$$A_\Phi=\sin\Phi\cot\theta \tag{34}$$

$$A_\psi=-\frac{\sin\Phi}{\sin\theta} \tag{35}$$

Thus

$$M_x=\hbar/i\left[\cos\Phi\frac{\partial}{\partial\theta}-\sin\Phi\cot\theta\frac{\partial}{\partial\Phi}+\frac{\sin\Phi}{\sin\theta}\frac{\partial}{\partial\psi}\right] \tag{36}$$

One can, of course, proceed in a completely analogous way to get the remaining components of the angular momentum, however, let us note from Eqs. (20) that

$$\frac{\partial\vartheta_1}{\partial\Phi}=\frac{\partial\vartheta_2}{\partial\Phi}=0 \tag{37}$$

and from (10) and (11)

$$\frac{\partial\varphi_1}{\partial\Phi}=\frac{\partial\varphi_2}{\partial\Phi}=1 \tag{38}$$

Since

$$\frac{\partial}{\partial\Phi}=\sum_{j=1}^{2}\left(\frac{\partial\varphi_j}{\partial\Phi}\cdot\frac{\partial}{\partial\varphi_j}+\frac{\partial\vartheta_j}{\partial\Phi}\cdot\frac{\partial}{\partial\vartheta_j}\right)$$

substitution of (37) and (38) yields

$$\frac{\partial}{\partial\Phi}=\frac{\partial}{\partial\varphi_1}+\frac{\partial}{\partial\varphi_2}$$

However, since

$$M_z=\frac{\hbar}{i}\left(\frac{\partial}{\partial\varphi_1}+\frac{\partial}{\partial\varphi_2}\right)$$

we therefore have the $z$-component of $\mathbf{M}$:

$$M_z=\frac{\hbar}{i}\frac{\partial}{\partial\Phi} \tag{39}$$

The remaining component of the angular momentum may be derived from the commutation relation $[M_z, M_x]=i\hbar M_y$
Straightforward substitution yields:

$$M_y=\frac{\hbar}{i}\left[\sin\Phi\frac{\partial}{\partial\theta}+\cos\Phi\cot\theta\frac{\partial}{\partial\Phi}-\frac{\cos\Phi}{\sin\theta}\cdot\frac{\partial}{\partial\psi}\right] \tag{40}$$

These relations are independent of $\theta_{12}$ corresponding to the statement that the angular momentum only depends on the (three) Euler angles $\theta$, $\Phi$, $\psi$. The forms of the three operators is the same as one gets with the Hylleraas-Breit angles.[3] The square of the angular momentum is likewise the same. One finds directly from the sum of the squares that

$$\mathbf{M}^2=-\hbar^2\left[\frac{\partial^2}{\partial\theta^2}+\frac{1}{\sin^2\theta}\left(\frac{\partial^2}{\partial\Phi^2}+\frac{\partial^2}{\partial\psi^2}\right)+\cot\theta\frac{\partial}{\partial\theta}-\frac{2\cot\theta}{\sin\theta}\frac{\partial^2}{\partial\Phi\partial\psi}\right] \tag{41}$$

The vector spherical harmonics, which have been given for the restricted range $m\geq k\geq 0$ in Eq. (25), are the simultaneous eigenfunctions of $\mathbf{M}^2$ with eigenvalue $\hbar^2\ell(\ell+1)$ and $\hbar m$

$$\mathbf{M}^2 D_\ell^{m,k}(\theta,\Phi,\psi)=\hbar^2\,\ell(\ell+1)\,D_\ell^{m,k}(\theta,\Phi,\psi)$$

$$\mathbf{M}_z D_\ell^{m,k}(\theta,\Phi,\psi)=\hbar m\,D_\ell^{m,k}(\theta,\Phi,\psi) \tag{43}$$

They are given in a completely general, normalized form in Pauling and Wilson.[10] With our phase [8] they are explicity

$$D_\ell^{m,k}(\theta,\Phi,\psi)=(-1)^{\frac{1}{2}[|k-m|+k-m]}N_{\ell mk}\sin^{|k-m|}\frac{\theta}{2}\cos^{|k+m|}\frac{\theta}{2}\times e^{i(m\Phi-k\psi)}F\left(-\ell+\frac{1}{2}\beta-1,\ \ell+\beta/2,\ 1+|k-m|,\ \sin^2\frac{\theta}{2}\right) \tag{44}$$

where

$$\beta=|k+m|+|k-m|+2 \tag{45}$$

and the normalization constant is

$$N_{\ell mk}=\left[\frac{(2\ell+1)(\ell+\frac{1}{2}|k+m|+\frac{1}{2}|k-m|)!(\ell-\frac{1}{2}|k+m|+\frac{1}{2}|k-m|)!}{8\pi_2(\ell-\frac{1}{2}|k+m|-\frac{1}{2}|k-m|)!(\ell+\frac{1}{2}|k+m|-\frac{1}{2}|k-m|)!}\right]^{\frac{1}{2}}\times\frac{1}{|k-m|!} \tag{46}$$

In addition to the usual magnetic quantum number $m$, the vector spherical harmonics depend on the quantum number $k$, an integer whose range of values is the same as $m$: $-\ell\leq m,\ k\leq\ell$.

The physical significance of $k$ derives from the fact that the $D_l^{m,k}$ are the eigenfunctions of the spherical top (for which (42) is the Schrödinger equation), and $k$ is the angular momentum quantum number about the body-fixed axis of rotation. With regard to the applications that we contemplate here, $k$ can be considered a degeneracy label which must be adjusted such that other requirements are fulfilled.

## V. CONSTRUCTION OF THE TOTAL WAVE FUNCTION

We shall confine ourselves here strictly to the atomic problem which implies that the potential energy as well as the kinetic energy commutes with the total angular momentum. In this case the total wave function for a given $l$ must be a linear superposition of the degenerate $D_l^{m,k}$. In addition, $m$ will be fixed for a given magnetic substate and the "radial" equations will be independent of m (cf. Appendix II).

Considering, for the moment, the residual coordinates as $r_1$, $r_2$, $\theta_{12}$, we can therefore expand the total wave function in the form:

$$\psi lm(\mathbf{r}_1, \mathbf{r}_2) = \sum_{k=-l}^{l} g^k(r_1, r_2, \theta_{12}) D_l^{m,k}(\theta, \Phi, \psi) \qquad (47)$$

The parity operations, Eq. (23), only effects the Euler angles, and from Eq. (28) it only multiplies the $D_l^{m,k}$ function by $(-1)^k$. Therefore by restricting the sum to even and odd values of $k$, we guarantee that the superpositions have even and odd parity respectively:

$$\Psi_{lm}^{\text{even}}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{k \text{ even}}'' g_l^k(r_1, r_2, \theta_{12}) D_l^{m,k}(\theta, \Phi, \psi) \qquad (48a)$$

$$\Psi_{lm}^{\text{odd}}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{k \text{ odd}}'' g_l^k(r_1, r_2, \theta_{12}) D_l^{m,k}(\theta, \Phi, \psi) \qquad (48b)$$

where the double prime on the summation emphasizes that the sum goes over every other value of $k$.

In deriving the radial equations (next section) we shall exploit the invariance of the radial equations with respect to $m$, by choosing $m=0$. When the Hamiltonian is written in terms of the Euler angles and the remaining variables, there will occur terms involving $\frac{\partial}{\partial\Phi}$ and $\frac{\partial}{\partial\psi}$. By virtue of $m=0$, the former terms vanish, but the latter terms would bring down the imaginery coefficient $ik$. In order to avoid complex equations, it is therefore convenient to construct real angular momentum functions. Let

$$D_l^{\varkappa+} = \frac{1}{\sqrt{2}} \frac{(D_l^{0,\varkappa} + D_l^{0,-\varkappa})}{(1+\delta_{0\varkappa}(\sqrt{2}-1))} = \frac{\sqrt{2l+1}}{2\pi} \frac{\cos \varkappa\psi}{(1+\delta_{0\varkappa}(\sqrt{2}-1))} d_l^{0,\varkappa}(\theta) \qquad (49a)$$

$$D_l^{\varkappa-} = \frac{1}{\sqrt{2}i}(D_l^{0,\varkappa} - D_l^{0,-\varkappa}) = \frac{\sqrt{2l+1}}{2\pi} \cdot \sin \varkappa\psi d_l^{0,\varkappa}(\theta) \qquad (49b)$$

for $\varkappa > 0$, where

$$\varkappa = |k|, \qquad (50)$$

The important property of these linear combinations is that they are eigenfunctions of exchange:

$$E_{12} D_l^{\varkappa\pm} = \pm(-1)^l D_l^{\varkappa\pm} \qquad (49c)$$

This then constitutes a set of real, orthonormal vector spherical harmonics. These real vector spherical harmonics are still eigenfunctions of parity with eigenvalue $(-1)^\varkappa$.

The property of exchange is a mite more complicated than parity in the sense that it affects not only the Euler angles, but the residual coordinates as well. The beauty and importance of Holmberg's angles, however, is that there is no mixing, and independent of whether we consider the residual variables $r_1$, $r_2$, $\theta_{12}$ or $r_1$, $r_2$, $r_{12}$, the effect of exchange on the residual coordinates is simply $r_1 \rightleftarrows r_2$.

Finally then, if we construct

$$\Psi_{l0}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{\varkappa}'' [f_l^{\varkappa+}(r_1, r_2, \theta_{12}) \cdot D_l^{\varkappa+}(\theta, \Phi, \psi) + f_l^{\varkappa-}(r_1, r_2, \theta_{12}) D_l^{\varkappa-}(\theta, \Phi, \psi)], \qquad (51)$$

the operation of exchange on this sum then gives, with the use of (49c),

$$E_{12}\psi_{l0}(\mathbf{r}_1, \mathbf{r}_2) = \psi_{l0}(\mathbf{r}_2, \mathbf{r}_1) = \sum_{\varkappa}'' [f_l^{\varkappa+}(\mathbf{r}_2, \mathbf{r}_1, \theta_{12})(-1)^l D_l^{\varkappa+}(\theta, \Phi, \psi) + f_l^{\varkappa-}(\mathbf{r}_2, \mathbf{r}_1, \theta_{12})(-1)^{l+1} D_l^{\varkappa-}(\theta, \Phi, \psi)]$$

Thus if

$$f_l^{\varkappa+}(r_2, r_1, \theta_{12}) = \pm(-1)^l f_l^{\varkappa+}(r_1, r_2, \theta_{12}) \quad (52a)$$

$$f_l^{\varkappa-}(r_2, r_1, \theta_{12}) = \pm(-1)^{l+1} f_l^{\varkappa-}(r_1, r_2, \theta_{12}), \quad (52b)$$

the function $\Psi_{\ell 0}$ of Eq. (51) is a real, space symmetric (upper sign) or space antisymmetric (lower sign), eigenfunction of $\mathbf{M}^2$ and $M_z$ corresponding to the quantum numbers $l$ and $m$ with $m=0$. The space symmetric and antisymmetric solutions correspond to singlet and triplet spin states respectively. Furthermore the restriction to $m=0$ is sufficient for deriving the radial equations.

We have shown that the $m=0$ function can be written in manifestly real form, Eq. (51). However, in that form, it is not obvious what the generalization is to arbitrary $m$ states. The generalization is nevertheless simply obtained.

Let

$$g_l^{\varkappa} = \frac{1}{\sqrt{2}}(f_l^{\varkappa+} - i f_l^{\varkappa-}) \quad \varkappa > 0 \quad (53)$$

and

$$g_l^{-\varkappa} = \frac{1}{\sqrt{2}}(f_l^{\varkappa+} + i f_l^{\varkappa-})$$

then the form (51) reduces to that of Eq. (47) for $m=0$. For arbitrary $m$ one then need only replace the $D_l^{0,k}$ functions by the appropriate $D_l^{m,k}$ functions, the radial $g_l^k$ functions remaining the same.

Alternatively one can define generalizations of the $D_l^{\varkappa\pm}$, Eq. (49), for arbitrary $m$.

$$D_l^{(m,\varkappa)+} \equiv [\sqrt{2} + \delta_{0\varkappa}(2-\sqrt{2})]^{-1}(D_l^{m,\varkappa} + D_l^{m,-\varkappa}) \quad m \neq 0$$

$$D_l^{(m,\varkappa)-} \equiv [\sqrt{2}i]^{-1}(D_l^{m,\varkappa} - D_l^{m,-\varkappa}) \quad (54)$$

with $D^{(0,\varkappa)\pm} \equiv D_l^{\varkappa\pm}$ for $m=0$. Note for $m \neq 0$ that the modified spherical harmonics, $D_l^{(m\varkappa)\pm}$ are no longer real.

The complete function for arbitrary $m$ can then be written

$$\psi_{lm}(\mathbf{r}_1, \mathbf{r}_2) = \sum' [f_l^{\varkappa+} D_l^{(m,\varkappa)+} + f_l^{\varkappa-} D_l^{(m,\varkappa)-}] \quad (55)$$

Note that by virtue of the action of the raising and lowering operators (Appendix II), the phase of the individual terms except possibly for an overall minus sign, is correct as it stands.

## VI. THE KINETIC ENERGY

Just as in the case of the angular momentum, the kinetic energy can be obtained by a straightforward process of partial differentiation. In this case, however, since second partial derivatives are involved, the differentiation is a much longer job, and, as we shall see, the partial derivatives involving $\theta_{12}$ no longer cancel out.

We start then with the kinetic energy in spherical coordinates

$$\nabla_1^2 + \nabla_2^2 = \frac{1}{r_1}\frac{\partial^2}{\partial r_1^2} r_1 + \frac{1}{r_2}\frac{\partial^2}{\partial r_2^2} r_2 + \frac{1}{r_1^2}\left[\frac{1}{\sin\vartheta_1}\frac{\partial}{\partial\theta_1}\sin\vartheta_1\frac{\partial}{\partial\vartheta_1} + \frac{1}{\sin^2\vartheta_1}\frac{\partial^2}{\partial\varphi_1^2}\right] + \frac{1}{r_2^2}\left[\frac{1}{\sin\vartheta_2}\frac{\partial}{\partial\vartheta_2}\sin\vartheta_2\frac{\partial}{\partial\vartheta_2} + \frac{1}{\sin^2\vartheta_2}\frac{\partial^2}{\partial\varphi_2^2}\right] \quad (56)$$

The first two terms are, of course, unaffected by the transformation. The angular differentiations then involve the transformation from the variables $\vartheta$, $\varphi_1$, $\vartheta_2$, $\varphi_2$ to $\theta$, $\Phi$, $\psi$ and $\theta_{12}$.

Consider the coefficient of the $r_1^{-2}$ term. After some regrouping, we can write

$$\frac{1}{\sin\vartheta_1}\frac{\partial}{\partial\vartheta_1}\sin\vartheta_1\frac{\partial}{\partial\vartheta_1} + \frac{1}{\sin^2\vartheta_1}\frac{\partial^2}{\partial\varphi^2} = \sum_{\alpha=1}^{\psi_1}\left[\left(\frac{\partial\varkappa_\alpha}{\partial\vartheta_1}\right)^2 + \frac{1}{\sin^2\vartheta_1}\left(\frac{\partial\varkappa_\alpha}{\partial\varphi_1}\right)^2\right]\frac{\partial^2}{\partial\varkappa_\alpha^2} + \sum_\alpha\left[\frac{\partial^2\varkappa_\alpha}{\partial\vartheta_1^2} + \frac{1}{\sin^2\vartheta_1}\frac{\partial^2\varkappa_\alpha}{\partial\varphi_1^2} + \cot\vartheta_1\frac{\partial\varkappa_\alpha}{\partial\vartheta_1}\right]\frac{\partial}{\partial\varkappa_\alpha} + 2\sum_{\alpha>}\sum_{\beta=1}\left[\frac{\partial\varkappa_\alpha}{\partial\vartheta_1}\frac{\partial\varkappa_\beta}{\partial\vartheta_1} + \frac{1}{\sin^2\vartheta_1}\frac{\partial\varkappa_\alpha}{\partial\varphi_1}\frac{\partial\varkappa_\beta}{\partial\varphi_1}\right]\frac{\partial^2}{\partial\varkappa_\alpha\partial\varkappa_\beta} \quad (57)$$

where for $\alpha=1, 2, 3, 4$, $\varkappa_\alpha$ refers to $\theta$, $\Phi$, $\psi$, $\theta_{12}$. The problem thus reduces to finding each of the square brackets separately in terms of the Euler angles and $\theta_{12}$. The results are given in Table 1.

The kinetic energy thus becomes

$$\nabla_1^2 + \nabla_2^2 = \left\{\frac{1}{r_1}\frac{\partial^2}{\partial r_1^2} r_1 + \frac{1}{r_2}\frac{\partial^2}{\partial r_2^2} r_2 + \left(\frac{1}{r_1^2} + \frac{1}{r_2^2}\right)\frac{1}{\sin\theta_{12}}\frac{\partial}{\partial\theta_{12}}\sin\theta_{12}\frac{\partial}{\partial\theta_{12}}\right\} + \frac{F_1}{r_1^2} + \frac{F_2}{r_2^2} \quad (58)$$

where

$$F_1=\frac{1}{\sin^2\theta_{12}}\Big[\sin^2\Big(\psi+\frac{1}{2}\theta_{12}\Big)\frac{\partial^2}{\partial\theta^2}+\cos^2\Big(\psi+\frac{1}{2}\theta_{12}\Big)\cot\theta\frac{\partial}{\partial\theta}+\cos^2\Big(\psi+\frac{1}{2}\theta_{12}\Big)\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\Phi^2}$$
$$+\sin(2\psi+\theta_{12})\frac{\cot\theta}{\sin\theta}\frac{\partial}{\partial\Phi}-\sin(2\psi+\theta_{12})\frac{1}{\sin\theta}\frac{\partial^2}{\partial\theta\partial\Phi}+\sin(2\psi+\theta_{12})\cot\theta\frac{\partial^2}{\partial\psi\partial\theta}$$
$$-2\cos^2\Big(\psi+\frac{1}{2}\theta_{12}\Big)\frac{\cot\theta}{\sin\theta}\frac{\partial^2}{\partial\psi\partial\Phi}\Big]-\frac{\partial^2}{\partial\psi\partial\theta_{12}}+A_1\frac{\partial^2}{\partial\psi^2}+B_1\frac{\partial}{\partial\psi} \quad (59)$$

and

$$A_1=\frac{1}{\psi}+\frac{\cot^2\theta}{\sin^2\theta_{12}}\cos^2\Big(\psi+\frac{1}{2}\theta_{12}\Big) \quad (60)$$

$$B_1=-\frac{\cos\psi\sin(\psi+\theta_{12})}{\sin^2\theta_{12}}-\frac{\cot^2\theta}{\sin^2\theta_{12}}\sin(2\psi+\theta_{12})$$
$$+\frac{\cos\theta_{12}}{2\sin\Big(\frac{\theta_{12}}{2}\Big)}\Bigg(1-\frac{1}{\cos\Big(\frac{1}{2}\theta_{12}\Big)}\Bigg) \quad (61)$$

The expressions for $F_2$, $A_2$ and $B_2$ can be obtained by replacing $\theta_{12}$ by $-\theta_{12}$ in the above formulae (including the appropriate partial derivatives). This is equivalent to operating with $E_{12}$.

$$F_2(\theta,\Phi,\psi,\theta_{12})=F_1(\theta,\Phi,\psi,-\theta_{12}) \quad (62)$$

It is clear, since all the coefficients are independent of $\Phi$, that $M_z$ commutes with the kinetic energy. We have also explicitly verified that $[M_z,\triangle 1^2+\triangle 2^2]=0$.

Note that the partial derivative involving $\theta_{12}$ and no other angles has been placed in the curly brackets with the radial derivatives. This is because this term, as the radial derivatives themselves, do not affect the orbital angular momentum, and are the terms which act on total $S$ states[1,2,3]

In fact; in the action of the remaining terms on the angular momentum eigenfunction rests the bulk of the reduction of the Schrödinger equation to its 3-dimensional "radial" form. With this reduction in mind (cf. next section), it is convenient to write $F_1$ in terms of operators whose effect on the angular momentum eigenfunctions is particularly simple. One can show

$$F_1=\frac{1}{2}\sin^2\theta_{12}\Big[-\frac{1}{\hbar^2}\mathbf{M}^2+\cos\theta_{12}(\sin 2\psi\Lambda_2$$
$$-\cos 2\psi\Lambda_1)+\sin\theta_{12}(\sin 2\psi\Lambda_1+\cos 2\psi\Lambda_2)\Big]$$
$$-\frac{\partial^2}{\partial\theta_{12}\partial\psi}+\Big(\frac{1}{\psi}-\frac{1}{2\sin^2\theta_{12}}\Big)\frac{\partial^2}{\partial\psi^2}$$
$$+\Bigg(\frac{\cos\theta_{12}}{2\sin\Big(\frac{1}{2}\theta_{12}\Big)}-\cot\theta_{12}-\frac{1}{2\sin\theta_{12}}\Bigg)\frac{\partial}{\partial\psi} \quad (63)$$

where

$$\Lambda_1=2\frac{\partial^2}{\partial\theta_2}+\frac{\partial^2}{\partial\psi^2}+\frac{1}{\hbar^2}\mathbf{M}^2$$
$$\Lambda_2=2\frac{\cot\theta}{\sin\theta}\frac{\partial}{\partial\Phi}-(1+2\cot^2\theta)\frac{\partial}{\partial\psi}$$
$$-\frac{2}{\sin\theta}\frac{\partial^2}{\partial\theta\partial\Phi}+2\cot\theta\frac{\partial^2}{\partial\theta\partial\psi} \quad (64)$$

and $\mathbf{M}^2$ is the total angular momentum squared operator given in Eq. (41). $F_2$ is again derivable from $F_1$ by replacing $\theta_{12}$ by $-\theta_{12}$.

## VII. THE REDUCED OR RADIAL EQUATIONS, ATOMIC CASE

The essential properties of the combinations of the operators appearing in $F_1$ and $F_2$, Eq. (62) are the following (cf. Appendix I):

$$(\sin 2\psi\Lambda_2-\cos 2\psi\Lambda_1)D_l^{\varkappa+}=-2\varkappa(\varkappa-1)B_{l\varkappa}(1-\delta_{0\varkappa})D_l^{(\varkappa-2)+}+\delta_{1\varkappa}(1-2A_\varkappa^l)D_l^{\varkappa+}$$
$$-\frac{(1+3\delta_{0\varkappa})}{(1+\delta_{0\varkappa}(\sqrt{2}-1))}\frac{A_\varkappa^l A_{\varkappa+1}^l}{2B_{l,\varkappa+2}}D_l^{(\varkappa+2)+} \quad (65a)$$

$$(\sin 2\psi\Lambda_1+\cos 2\psi\Lambda_2)D_l^{\varkappa+}=-2\varkappa(\varkappa-1)B_{l\varkappa}(1-\delta_{2\varkappa})(1-\delta_{0\varkappa})D_l^{(\varkappa-2)-}-\delta_{1\varkappa}(1-2A_\varkappa^l)D_l^{\varkappa-}$$
$$+\frac{1+3\delta_{0\varkappa}}{(1+\delta_{0\varkappa}(\sqrt{2}-1))}\frac{A_\varkappa^l A_{\varkappa+1}^l}{2B_{l,\varkappa+2}}D_l^{(\varkappa+2)-} \quad (65b)$$

$$(\sin 2\psi\Lambda_2-\cos 2\psi\Lambda_1)D_\ell^{\varkappa-}=-2\varkappa(\varkappa-1)B_{\ell\varkappa}(1-\delta_{2\varkappa})(1-\delta_{0\varkappa})D_\ell^{(\varkappa-2)-}-\delta_{1\varkappa}(1-2A_\varkappa^\ell)D_\ell^{\varkappa-}$$
$$-\frac{(1-\delta_{0\varkappa})}{(1+\delta_{0\chi}(\sqrt{2}-1))}\frac{A_\varkappa^\ell A_{\varkappa+1}^\ell}{2B_{\ell,\varkappa+2}}D_\ell^{(\varkappa+2)-} \quad (66a)$$

$$(\sin 2\psi\Lambda_1+\cos 2\psi\Lambda_2)D_\ell^{\varkappa-}=2\varkappa(\varkappa-1)B_{\ell\varkappa}(1-\delta_{0\varkappa})D_\ell^{(\varkappa-2)+}-\delta_{1\varkappa}(1-2A_\varkappa^\ell)D_\ell^{\varkappa+}$$
$$-\frac{(1-\delta_{0\varkappa})}{(1+\delta_{0\varkappa}(\sqrt{2}-1))}\frac{A_\varkappa^\ell A_{\varkappa+1}^\ell}{2B_{\ell\,\varkappa+2}}D_\ell^{(\varkappa+2)+} \quad (66b)$$

where

$$A_\varkappa^\ell=\frac{(\varkappa-\ell)(\varkappa+\ell+1)}{2(\varkappa+1)} \quad (67)$$

and

$$B_{\ell\varkappa}=\frac{[(\ell-\varkappa+1)(\ell-\varkappa+2)(\ell+\varkappa)(\ell+\varkappa-1)]^{\frac{1}{2}}}{4\varkappa(\varkappa-1)} \quad (68)$$

Recall that $\varkappa$ is the absolute value of $k$, Eq. (50).

We have proved in Appendix II that these relationships are not altered if one replaces the $D_\ell^{\varkappa\pm}$ functions by the $D_\ell^{(m,\varkappa)\pm}$ functions of Eq. (54). As such it becomes quite simple to derive the reduced equations from the original Schrödinger equation

$$H\psi_{\ell m}=E\psi_{\ell m} \quad (69)$$

for any $m$, and to see that the radial equations are independent of $m$. One obtains

$$\left[L_{\theta 12}+\frac{2m}{\hbar^2}(E-V)\right]f_\ell^{\varkappa+}-\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\left[\left\{\frac{\ell(\ell+1)-\varkappa^2}{2\sin^2\theta_{12}}+\frac{\varkappa^2}{4}-\frac{\cot\theta_{12}}{2\sin\theta_{12}}\delta_{1\varkappa}(1-2A_\varkappa^\ell)\right\}f_\ell^{\varkappa+}\right.$$
$$+\frac{\cot\theta_{12}}{\sin\theta_{12}}(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}f_\ell^{(\varkappa+2)+}+\frac{\cot\theta_{12}}{4\sin\theta_{12}}\frac{(1+3\delta_{0,\varkappa-2})}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-1}^\ell A_{\varkappa-2}^\ell}{B_{\ell\varkappa}}f_\ell^{(\varkappa-2)+}\Big]$$
$$+\left(\frac{1}{r_1^2}-\frac{1}{r_2^2}\right)\left[\varkappa\left(\frac{\cos\theta_{12}}{2\sin\theta_{12}}-\cot\theta_{12}-\frac{1}{2\sin\theta_{12}}\right)f_\ell^{\varkappa-}-\varkappa\frac{\partial f_\ell^{\varkappa-}}{\partial\theta_{12}}-\frac{\delta_{1\varkappa}(1-2A_\varkappa^\ell)}{2\sin\theta_{12}}f_\ell^{\varkappa-}\right.$$
$$\left.+\frac{(\varkappa+1)(\varkappa+2)}{\sin\theta_{12}}B_{\ell,\varkappa+2}f_\ell^{(\varkappa+2)-}-\frac{1}{4\sin\theta_{12}}\frac{(1-\delta_{0,\varkappa-2})}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{B_{\ell\varkappa}}f_\ell^{(\varkappa-2)-}\right]=0 \quad (70a)$$

$$\left[L_{\theta 12}+\frac{2m}{\hbar^2}(E-V)\right]f_\ell^{\varkappa-}-\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\left[\left\{\frac{\ell(\ell+1)-\varkappa^2}{2\sin^2\theta_{12}}+\frac{\varkappa^2}{4}+\frac{\cot\theta_{12}}{2\sin\theta_{12}}\delta_{1\varkappa}(1-2A_\varkappa^\ell)\right\}f_\ell^{\varkappa-}\right.$$
$$+\frac{\cot\theta_{12}}{\sin\theta_{12}}(1-\delta_{2,\varkappa+2})(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}f_\ell^{(\varkappa+2)-}+\frac{\cot\theta_{12}}{4\sin\theta_{12}}\frac{(1-\delta_{0,\varkappa-2})}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{B_{\ell\varkappa}}f_\ell^{(\varkappa-2)-}\Big]$$
$$+\left(\frac{1}{r_1^2}-\frac{1}{r_2^2}\right)\left[-\varkappa\left(\frac{\cos\theta_{12}}{2\sin\left(\frac{\theta_{12}}{2}\right)}-\cot\theta_{12}-\frac{1}{2\sin\theta_{12}}\right)f_\ell^{\varkappa+}+\varkappa\frac{\partial f_\ell^{\varkappa+}}{\partial\theta_{12}}-\frac{\delta_{1\varkappa}(1-2A_\varkappa^\ell)}{2\sin\theta_{12}}f_\ell^{\varkappa+}\right.$$
$$\left.-\frac{(1-\delta_{2,\varkappa+2})}{\sin\theta_{12}}(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}f_\ell^{(\varkappa+2)+}+\frac{1}{4\sin\theta_{12}}\frac{(1+3\delta_{0,\varkappa-2})}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{B_{\ell\varkappa}}f_\ell^{(\varkappa-2)+}\right]=0 \quad (70b)$$

$L_{\theta 12}$ is the $S$ wave part of the kinetic energy, and only the term containing it survives in the description of $S$ states:[1,2,3]

$$L_{\theta 12}=\frac{1}{r_1}\frac{\partial^2}{\partial r_1^2}r_1+\frac{1}{r_2}\frac{\partial^2}{\partial r_2^2}r_2$$
$$+\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\frac{1}{\sin\theta_{12}}\frac{\partial}{\partial\theta_{12}}\left(\sin\theta_{12}\frac{\partial}{\partial\theta_{12}}\right) \quad (71)$$

Eqs. (70) are the "radial" equations, which it has been our purpose to derive. They pertain to both types of parity and exchange states. Parity is determined by the evenness or oddness of $\varkappa$. If, for example, $\ell$ is even, and we want to describe a state of even parity, Eqs. (70) couple the functions $f_\ell^{\varkappa+}$ and $f_\ell^{\varkappa-}$ for $\varkappa=0,2,4\ldots\ell$. This involves $\ell/2$ pairs plus one function (for $\varkappa=0$, $D_\ell^{0-}$ is zero hence $f_\ell^{0-}$ can be taken to be zero)

or $\ell+1$ functions. The odd parity equations for the same $\ell$ correspond to the coupling of the function with $\varkappa=1,3,\ldots,\ell-1$. This relates $\ell/2$ pairs or $\ell$ functions to each other. Both even and odd parity together therefore involve $(2\ell+1)$ functions corresponding to the $(2\ell+1)$ degeneracy of the vector spherical harmonics for a given $m$. For $\ell$ odd, there are $\ell$ functions involved in the even parity equations and $\ell+1$ functions in the odd parity equations.

*For a given parity and $\ell$, both singlet and triplet (space symmetric and antisymmetric) states are described by the same set of equations.* The differences in the solutions devolve from the different boundary conditions which must be applied, Eqs. (52). One of the key virtues of the functions $f_\ell^{\varkappa\pm}(r_1, r_2, \theta_{12})$ is that they are either symmetric or antisymmetric; thus they may be confined to the region, say, $r_1\geq r_2$. If, for example, the exchange character of $f_\ell^{\varkappa+}$ is symmetric (which, according to (52), implies that $f_\ell^{\varkappa-}$ is antisymmetric), then these properties may be embodied in the boundary conditions [11]:

$$\left[\frac{\partial}{\partial n} f_\ell^{\varkappa+}(r_1, r_2, \theta_{12})\right]_{r_1=r_2}=0 \tag{72}$$

where $\frac{\partial}{\partial n}$ represents the normal derivative, and

$$[f_\ell^{\varkappa-}(r_1, r_2, \theta_{12})]_{r_1=r_2}=0 \tag{73}$$

and the solution from there on involves only the region $r_1>r_2\geq 0$. Such equations have distinct advantages from the point of view of numerical solutions.[12]

One can define, however, an asymmetric function in terms of which the radial equations can be more simply written. Letting

$$F_\ell^{\varkappa}(r_1,r_2,\theta_{12})\equiv f_\ell^{\varkappa+}(r_1,r_2,\theta_{12})+f_\ell^{\varkappa-}(r_1,r_2,\theta_{12}) \tag{74}$$

and

$$\tilde{F}_\ell^{\varkappa}(r_1, r_2, \theta_{12})\equiv f_\ell^{\varkappa+}(r_1, r_2, \theta_{12})-f_\ell^{\varkappa-}(r_1, r_2, \theta_{12}) \tag{75}$$

so that, from eq. (74),

$$\tilde{F}_\ell^{\varkappa}(r_2, r_1, \theta_{12})=\pm(-1)^\ell F_\ell^{\varkappa}(r_1, r_2, \theta_{12}) \tag{76}$$

for singlet (upper sign) and triplet (lower sign) cases.

one obtains by adding (70a) and (70b):

$$\begin{aligned}
&\left[L_{\theta 12}+\frac{2m}{\hbar^2}(E-V)\right]F_\ell^{\varkappa}-\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\left[\left(\frac{\ell(\ell+1)-\varkappa^2}{2\sin^2\theta_{12}}+\frac{\varkappa^2}{4}\right)F_\ell^{\varkappa}-\frac{\cot\theta_{12}}{2\sin\theta_{12}}\delta_{1\varkappa}(1-2A_\varkappa^\ell)\tilde{F}_\ell^{\varkappa}\right.\\
&+\frac{\cot\theta_{12}}{\sin\theta_{12}}(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}\left\{F_\ell^{\varkappa+2}-\frac{\delta_{2,\varkappa+2}}{2}(F_\ell^{\varkappa+2}-\tilde{F}_\ell^{\varkappa+2})\right\}+\frac{\cot\theta_{12}}{4\sin\theta_{12}}\frac{1}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1)}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{B_{\ell\varkappa}}\{F_\ell^{\varkappa-2}\\
&\left.+\delta_{0,\varkappa-2}(2\tilde{F}_\ell^{\varkappa-2}+F_\ell^{\varkappa-2})\}\right]+\left(\frac{1}{r_1^2}-\frac{1}{r_2^2}\right)\left[-\varkappa\left(\frac{\cos\theta_{12}}{2\sin\left(\frac{\theta_{12}}{2}\right)}-\cot\theta_{12}-\frac{1}{2\sin\theta_{12}}\right)\tilde{F}_\ell^{\varkappa}+\varkappa\frac{\partial\tilde{F}_\ell^{\varkappa}}{\partial\theta_{12}}\right.\\
&-\frac{\delta_{1\varkappa}(1-2A_\varkappa^\ell)}{2\sin\theta_{12}}F_\ell^{\varkappa}+\frac{(\varkappa+1)(\varkappa+2)}{\sin\theta_{12}}B_{\ell,\varkappa+2}\left\{-\tilde{F}_\ell^{\varkappa+2}+\frac{1}{2}\delta_{2,\varkappa+2}(F_\ell^{\varkappa+2}+\tilde{F}_\ell^{\varkappa+2})\right\}\\
&\left.\left.+\frac{1}{4\sin\theta_{12}}\frac{1}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{B_{\ell\varkappa}}\left\{\tilde{F}_\ell^{\varkappa-2}+\delta_{0,\varkappa-2}(2F_\ell^{\varkappa-2}+\tilde{F}_\ell^{\varkappa-2})\right\}\right]=0
\end{aligned} \tag{77}$$

These equations, depending as they do on $F_\ell$ and $\tilde{F}_\ell$, are more analogous to the form of the P-wave equation of Breit.[2] The question may arise in connection with these as well as Breit's equations, of whether they are well-defined, since they involve two functions $F_\ell^{\varkappa}$ and $F_\ell^{\varkappa}$ and yet there is only one equation (for a given $\varkappa$). This question, in fact would appear to be particularly relevant as the previous form of our equations, (70) do constitute a coupled set for a given $\varkappa$. To see that both situations are meaningful and in particular that (77) is well-defined, consider a numerical solution of (77). In that case the space of the independent variables is divided into a grid of points, and $F_\ell^{\varkappa}$ is the collection of numbers associated with these grid points. $F_\ell^{\varkappa}$ can

therefore be considered a vector with as many components as there are grid points. The differential equation is replaced by a matrix which operates on the vector $F_l^\varkappa$. Now everytime an $\tilde{F}_l^\varkappa$ occurs in the equation, it is completely clear what has to be done: namely one must let the matrix counterpart of its coefficient in the differential equation operate on that component of $F_l^\varkappa$ which is its reflection point defined by (76). This is a completely unambiguous prescription which is tantamount to saying that the set (77) is well-defined by itself. The reason that (70) is composed of two equations for each $\varkappa$ whereas (77) is not is due to the fact that the functions $F_l^\varkappa$ are asymmetric and therefore must be solved for in the whole $r_1$, $r_2$, $\theta_{12}$ space. On the other hand the $f_l^{\varkappa\pm}$ functions are either symmetric or antisymmetric, and therefore they are restricted to the $r_1 \geq r_2$, $\theta_{12}$ (or equivalently to the $r_1 \leq r_2$, $\theta_{12}$) space. Since this is only half the independent variable space, it is necessary that there be double the number of functions to recover the same information. This is again to say that (70) and (77) are completely equivalent. (Nevertheless a redundant equation with $F_l^\varkappa$ and $\tilde{F}_l^\varkappa$ interchanged may readily be derived.)

We have stated that (70) has certain advantages from the point of view of numerical integration. However, it should also be stated that the form (77) will probably be more advantageous for ordinary variational calculations. This is because if one adopts a specific analytic form of $F_l^\varkappa$, one need only interchange $r_1$ and $r_2$ in the expression to obtain $\tilde{F}_l^\varkappa$.

The differences in the $f_l^{\varkappa\pm}$ description from that of $F_l^\varkappa$ gives rise to characteristic differences in the formulation of boundary conditions for scattering problems (cf. the next section).

The restriction of these equations to the atomic case (two identical particles in a fixed central field) has implicitly been made by assuming that the potential is a function of the residual coordinates,

$$V = V(r_1, r_2, r_{12}), \tag{78}$$

so that $V$ commutes with the angular momentum and therefore appears as an additional diagonal term in the radial equations. The inter-particle distance $r_{12}$ is related to the independent radial coordinates that we have thus far considered, $r_1$, $r_2$, $\theta_{12}$ via the law of cosines:

$$r_{12}^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos \theta_{12}.$$

Alternatively, however, one can consider $r_1$, $r_2$, and $r_{12}$ as the independent coordinates and derive radial equations involving them. Those coordinates, in fact, have certain advantages since the three singularities in the potential occur at their null points. As such they can describe the wave function in the region of close interaction very well. These variables, therefore, are particularly suited to calculation of lowlying bound states of two-electron atoms (where on the whole the electrons are quite close to each other and to the nucleus) and such successful calculations have been done ever since the early work of Hylleraas.[1]

When one considers the equation in the form we have previously given them, involving $\theta_{12}$, one is naturally led to expand the "radial" wave function in terms of Legendre polynomials of cos $\theta_{12}$[11]. The expansion is then truncated after some $P_n$ (cos $\theta_{12}$) and convergence is sought as a function of $n$. In these classes of two electron problems, this constitutes the idea of configuration interaction in its most general form. Recently this idea has come under some criticism [13,14,15] principally because such a relative partial wave expansion necessarily converges slowly where the electron-electron interaction is large ($r_{12}$ small). The argument has some validity for the above-mentioned low-lying bound states. However, the argument can easily get distorted and exaggerated, for instance when applied to the low-energy scattering of electrons from hydrogen.[14] The point there is that the long-range correlation coming from the induced potential in the atom is at least as important as the short range correlations [16] and yet is only poorly approximated by the conventional Hylleraas type of expansion. This situation has been discussed elsewhere.[12]

These reservations notwithstanding, however, it is nevertheless true that the most accurate three body calculations have been made using the $r_1$, $r_2$, $r_{12}$ coordinates, or linear combinations of them,[17] on the low-lying states of helium [18] and its isoelectronic ions.[17] We therefore give below the radial equations in terms of $r_1$, $r_2$, $r_{12}$. The equations are in their asymmetric form corres-

ponding to Eq. (77), since it is assumed that they will be utilized in connection with variational calculations with analytic expansions of the radial wave functions.

$$\left[L_{r12}+\frac{2m}{\hbar^2}(E-V)\right]F_\ell^\varkappa-\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\Bigg[\left\{\frac{(\ell(\ell+1)-\varkappa^2)}{\zeta^2}2r_1^2r_2^2+\frac{\varkappa^2}{4}\right\}F_\ell^\varkappa-\delta_{1\varkappa}(1-2A_\varkappa^\ell)(r_1^2+r_2^2-r_{12}^2)\frac{r_1r_2}{\zeta^2}\tilde{F}_\ell^\varkappa$$

$$+(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}(r_1^2+r_2^2-r_{12}^2)\frac{2r_1r_2}{\zeta^2}\left\{F_\ell^{\varkappa+2}-\frac{\delta_{2,\varkappa+2}}{2}(F_\ell^{\varkappa+2}-\tilde{F}_\ell^{\varkappa+2})\right\}$$

$$+\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell(r_1^2+r_2^2-r_{12}^2)}{2B_{\ell\varkappa}[1+\delta_{0,\varkappa-2}(\sqrt{2}-1)]}\frac{r_1r_2}{\zeta^2}\{F_\ell^{\varkappa-2}+\delta_{0,\varkappa-2}(2\tilde{F}_\ell^{\varkappa-2}+F_\ell^{\varkappa-2})\}\Bigg]+\left(\frac{1}{r_1^2}-\frac{1}{r_2^2}\right)\Bigg[-\varkappa\Bigg\{\frac{1}{2}\frac{r_1^2+r_2^2-r_{12}^2}{\sqrt{r_1r_2[r_{12}^2-(r_1-r_2)^2]}}$$

$$-\frac{r_1^2+r_2^2-r_{12}^2+r_1r_2}{\zeta}\Bigg\}\tilde{F}_\ell^\varkappa+\frac{\varkappa\zeta}{2r_{12}}\frac{\partial\tilde{F}_\ell^\varkappa}{\partial r_{12}}+(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}\frac{2r_1r_2}{\zeta}\left\{-\tilde{F}_\ell^{\varkappa+2}+\frac{1}{2}\delta_{2,\varkappa+2}(F_\ell^{\varkappa+2}+\tilde{F}_\ell^{\varkappa+2})\right\}$$

$$+\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{2B_{\ell\varkappa}}\frac{r_1r_2}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{1}{\zeta}\{\tilde{F}_\ell^{\varkappa-2}+\delta_{0,\varkappa-2}(2F_\ell^{\varkappa-2}+\tilde{F}_\ell^{\varkappa-2})\}-\delta_{1\varkappa}(1-2A_\varkappa^\ell)\frac{r_1r_2}{\zeta}\tilde{F}_\ell^\varkappa\Bigg]=0 \quad (79)$$

Here

$$\zeta=+[r_{12}^4-(r_1^2-r_2^2)^2+2r_{12}^2(r_1^2+r_2^2)]^{\frac{1}{2}} \quad (80)$$

The quantity whose square root $\zeta$ is can easily be shown to be positive definite. In the equation (79) the $F_\ell^\varkappa$ is understood to be a function of $r_1, r_2, r_{12}$:

$$F_\ell^\varkappa=F_\ell^\varkappa(r_1, r_2, r_{12})$$

In addition $L_{r12}$ is the kinetic energy counterpart of the $S$-wave $L_{\theta12}$ in terms of $r_1, r_2, r_{12}$:

$$L_{r12}=\frac{1}{r_1}\frac{\partial^2}{\partial r_1^2}r_1+\frac{1}{r_2}\frac{\partial^2}{\partial r_2^2}r_2+\frac{2}{r_{12}}\frac{\partial^2}{\partial r_{12}^2}r_{12}$$

$$+(r_1^2+r_{12}^2-r_2^2)\frac{1}{r_1r_2}\frac{\partial^2}{\partial r_1\partial r_{12}}$$

$$+(r_2^2+r_{12}^2-r_1^2)\frac{1}{r_1r_2}\frac{\partial^2}{\partial r_2\partial r_{12}} \quad (81)$$

The equations (79) can readily be put in the form of coupled equations for a given $\varkappa$. In that form they would be closest to the form originally given by Holmberg,[6] (although as we have stated his equations apply to three equal mass particles). One salient difference between the two sets of equations, however, is that the present ones are manifestly real, whereas one term in Holmberg's equations is imaginary.[6] It is clear that the equations as well as the solutions must be reducible to completely real form for any given angular momentum state. The accomplishment of this in the present case comes from the explicit construction of real vector spherical harmonics, Eq. (49).

## VIII. BOUNDARY CONDITIONS FOR SCATTERING

In this section we derive the asymptotic forms of the radial functions corresponding to the scattering of an electron from a one electron atom in its ground state. The Coulomb modifications when the target system is an ion instead of an atom can readily be made and will have no effect on the angular integrations with which we are here concerned.

As we have seen in the foregoing sections culminating in the last section, the selection of a symmetric choice of Euler angles (Holmberg's angles) has allowed for a completely general derivation of the radial equations. From the point of view of a scattering problem, however, a symmetric choice of angles is not the most advantageous since here we are concerned with an intrinsically asymmetric situation. Thus if we consider that region of configuration space where $r_1$ is large and $r_2$ small, corresponding to electron 1 being scattered from the atom to which electron 2 is bound, the wave function *in this region alone* will not be symmetric. However, in terms of the Hylleraas-Breit angles, the spherical angles of one of the particles being defined as two of the Euler angles. the wave function in this asymmetric region is easier to describe. Nevertheless this is a complication of detail only, since all the

angular integrations may readily be performed as we shall now show.

We start with the statement that the complete wave function must have the asymptotic form:

$$\lim_{r_1\to\infty}\psi_{\ell 0}(\mathbf{r}_1,\mathbf{r}_2)=\frac{1}{r_1}\sin\left(kr_1+\delta\ell-\frac{\ell\pi}{2}\right)Y_{\ell 0}(\Omega_1)\frac{R_{1s}}{r_2}(r_2)Y_{00}(\Omega_2) \quad (82)$$

where $(R_{1s}(r_2)/r_2)\,Y_{00}(\Omega_2)$ is the ground state of the one-electron atom (hydrogen). On the other hand, from Eq. (51).

$$\lim_{r_1\to\infty}\psi_{\ell 0}(\mathbf{r}_1,\mathbf{r}_2)=\frac{1}{r_1}\sin\left(kr_1+\delta\ell-\frac{\ell\pi}{2}\right)\frac{R_{1s}(r_2)}{r_2}\times$$
$$\times\sum_{\varkappa}''[\alpha_\ell^{\varkappa+}(\theta_{12})D_\ell^{\varkappa+}(\theta,\Phi,\psi)+\alpha_\ell^{\varkappa-}(\theta_{12})D_\ell^{\varkappa-}(\theta,\Phi,\psi)] \quad (83)$$

where

$$\alpha_\ell^{\varkappa\pm}(\theta_{12})=Y_{00}\int_0^{\pi}\int_0^{2\pi}\int_0^{2\pi}D_\ell^{\varkappa\pm}(\theta,\Phi,\psi)Y_{\ell 0}(\vartheta_1)\sin\theta\,d\theta\,d\Phi\,d\psi \quad (84)$$

It should be noted that (82) refers to the state of parity $(-1)^\ell$ as long as we are considering elastic scattering from the ground (1s) state. This then defines the evenness of oddness of the values over which $\varkappa$ goes in the summation in Eq. (83).

The quadrature in (84) can readily be performed by recalling from Section II that $\vartheta_1$ is the angle between $\hat{\mathbf{z}}$ and $\hat{\mathbf{r}}_1$ whose spherical angles in the primed coordinate systems are given in Eq. (20). One can then use these spherical angles to expand $P_\ell(\cos\vartheta_1)$ via the addition theorems for spherical harmonics.

In its real form [9] this gives in the present case

$$P_\ell(\vartheta_1)=P_\ell(\pi/2)P_\ell(0)+2\sum_{m=1}^{\ell}(-1)^m\frac{(\ell-m)!}{(\ell+m)!}P_\ell^m(\pi/2)P_\ell^m(\theta)\cos m\left(\psi-\frac{1}{2}\theta_{12}\right) \quad (85)$$

In (85) we have written both the Legendre and associated Legendre polynomials as functions of the angle but what we mean in all cases is that the angle is to be substituted into the transcendental form of the function. For example $P_1(\beta)=\cos\beta$ and *not* $P_1(\beta)=\beta$. The sign of the $P_\ell^m$ is that of Magnus and Oberhettinger [9] (which differs by $(-1)^m$ from that of Morse and Feshbach [3]). To complete the quadrature in (84) we note that

$$d_\ell^{0,\varkappa}(\theta)=\left[\frac{(\ell-\varkappa)!}{(\ell+\varkappa)!}\right]^{\frac{1}{2}}P_\ell^\varkappa(\theta) \quad (86)$$

Substitution into (84) now yields the desired result:

$$\alpha_\ell^{\varkappa+}(\theta_{12})=(-1)^\varkappa P_\ell^\varkappa\left(\frac{\pi}{2}\right)\left[\frac{(\ell-\varkappa)!}{(\ell+\varkappa)!}\right]^{\frac{1}{2}}\left\{\frac{\delta_{0\varkappa}}{\sqrt{2}}+(1-\delta_{0\varkappa})\cos\left(\varkappa\frac{\theta}{2}12\right)\right\} \quad (87a)$$

$$\alpha_\ell^{\varkappa-}(\theta_{12})=(-1)^\varkappa P_\ell^\varkappa\left(\frac{\pi}{2}\right)\left[\frac{(\ell-\varkappa)!}{(\ell+\varkappa)!}\right]^{\frac{1}{2}}\sin\left(\varkappa\frac{\theta}{2}12\right) \quad (87b)$$

The radial function themselves thus approach

$$\lim_{r_1\to\infty}f_\ell^{\varkappa\pm}(r_1,r_2,\theta_{12})=\frac{1}{r_1}\sin\left(kr_1+\delta_\ell-\ell\frac{\pi}{2}\right)\frac{R_{1s}(r_2)}{r_2}\alpha_\ell^{\varkappa\pm}(\theta_{12}) \quad (88)$$

in which form we see that $r_1$, $r_2$ dependence of all the $f_\ell^{\varkappa\pm}$ functions is independent of $\varkappa$, so that none of them vanishes in the asymptotic region. Since in all cases the $\theta_{12}$ dependence is trivial, it may be worthwhile to define new functions whose asymptotic behaviour is strictly the $r_1$, $r_2$ dependence in (88).

For bound state problems, it is clear that all the radial functions must vanish in all asymptotic regions.

## IX. OTHER APPLICATIONS

In addition to two electron atomic or ionic systems the present equations apply to double *mu* or *pi* mesic atoms, although as the mass of the identical particles get heavier, the correction for the center of mass becomes more important. Also for the spinless bosons (*pi* mesons) only the space symmetric solutions will presumably be relevant.

The equations can also be applied to two different particles of the same mass (positron-hydrogen scattering, for example). In this case, the potential, $V$ will no longer be symmetric hence the solutions will not be symmetric which implies that boundary conditions like (72) must be

changed to matching conditions of the asymmetric solutions along the line $r_1 = r_2$.[19] This has the effect of giving one solution where formerly there were two, in accord with the distinguishability of the particles.

The major further application of this approach is to two-electron diatomic molecules. In this case, the extension from one [20] to two electrons is non-trivial. However, the analysis has been completed and will be published elsewhere.[21]

TABLE 1.—*Coefficients of the Angular Derivatives in the Kinetic Energy.*[a] *Cf. Eq. (49)*

| Coefficient | Derivative | Coefficient | Derivative |
|---|---|---|---|
| $\dfrac{\sin^2\left(\psi \mp \frac{\theta_{12}}{2}\right)}{\sin^2 \theta_{12}}$ | $\dfrac{\partial^2}{\partial \theta^2}$ | $\cot \theta_{12}$ | $\dfrac{\partial}{\partial \theta_{12}}$ |
| $\dfrac{\cos^2\left(\psi + \frac{\theta_{12}}{2}\right)}{\sin^2 \theta_{12} \sin^2 \theta}$ | $\dfrac{\partial^2}{\partial \Phi^2}$ | $\dfrac{-\sin (2\psi + \theta_{12})}{\sin^2 \theta_{12} \sin \theta}$ | $\dfrac{\partial^2}{\partial \theta \partial \Phi}$ |
| $A_1$ | $\dfrac{\partial^2}{\partial \psi^2}$ | $\dfrac{\cot \theta \sin (2\psi + \theta_{12})}{\sin_2 \theta_{12}}$ | $\dfrac{\partial^2}{\partial \theta \partial \psi}$ |
| 1 | $\dfrac{\partial^2}{\partial \theta_{12}^2}$ | 0 | $\dfrac{\partial^2}{\partial \theta \partial \theta_{12}}$ |
| $\dfrac{\cot \theta \cos^2\left(\psi + \frac{\theta_{12}}{2}\right)}{\sin^2 \theta_{12}}$ | $\dfrac{\partial}{\partial \theta}$ | $\dfrac{-2 \cos \theta \cos^2\left(\psi + \frac{\theta_{12}}{2}\right)}{\sin^2 \theta_{12} \sin^2 \theta}$ | $\dfrac{\partial^2}{\partial \Phi \partial \psi}$ |
| $\dfrac{\cos \theta \sin (2\psi + \theta_{12})}{\sin^2 \theta \sin^2 \theta^{12}}$ | $\dfrac{\partial}{\partial \Phi}$ | 0 | $\dfrac{\partial^2}{\partial \Phi \partial \theta_{12}}$ |
| $B_1$ | $\dfrac{\partial}{\partial \psi}$ | $-1$ | $\dfrac{\partial^2}{\partial \psi \partial \theta_{12}}$ |

[a] $A_1$ and $B_1$ are given in Eqs. (60) and (61).

# APPENDIX I

In this appendix, we prove the Eqs. (65) and (66). For $m=0$, $\frac{\partial}{\partial\Phi}$ terms give zero. Therefore, we can write

$$\Lambda_1 \rightarrow \frac{\partial^2}{\partial\theta^2} - \cot\theta\frac{\partial}{\partial\theta} - \cot^2\theta\frac{\partial^2}{\partial\psi^2} \tag{I-1}$$

$$\Lambda_2 \rightarrow 2\cot\theta\frac{\partial^2}{\partial\psi\partial\theta} - \frac{\partial}{\partial\psi} - 2\cot^2\theta\frac{\partial}{\partial\psi} \tag{I-2}$$

and

$$D_\ell^{\varkappa+} = N_{\ell\varkappa}\cos\varkappa\psi\sin^\varkappa\theta F_\varkappa \tag{I-3}$$

$$D_\ell^{\varkappa-} = N_{\ell\varkappa}\sin\varkappa\psi\sin^\varkappa\theta F_\varkappa \tag{I-4}$$

where

$$N_{\ell\varkappa} = \sqrt{\frac{2\ell+1}{8\pi^2}\frac{(\ell+\varkappa)!}{(\ell-\varkappa)!}}\frac{1}{\varkappa!}\frac{1}{2^\varkappa} \tag{I-5}$$

and

$$F_\varkappa = F\left(\varkappa-\ell,\ \varkappa+\ell+1,\ 1+\varkappa,\ \sin^2\frac{\theta}{2}\right) \tag{I-6}$$

Now

$$\Lambda_1 D_\ell^{\varkappa+} = N_{\ell\varkappa}\cos\varkappa\psi[2\varkappa(\varkappa-1)\ \sin^{\varkappa-2}\theta\cos^2\theta F_\varkappa - \varkappa\sin^\varkappa\theta F_\varkappa + 2\varkappa\sin^\varkappa\theta\cos\theta A_\varkappa^\ell F_{\varkappa+1} + \sin^{\varkappa+2}\theta A_\varkappa^\ell A_{\varkappa+1}^\ell F_{\varkappa+2}] \tag{I-7}$$

where we have used the well known relations for the derivatives of hypergeometric functions:

$$\frac{dF_\varkappa}{d\theta} = A_\varkappa^\ell\sin\theta F_{\varkappa+1} \tag{I-8}$$

$$\frac{d^2}{d\theta^2}F_\varkappa = A_\varkappa^\ell\cos\theta F_{\varkappa+1} + A_\varkappa^\ell A_{\varkappa+1}^\ell\sin^2\theta F_{\varkappa+2} \tag{I-9}$$

A relation between $F_\varkappa$, $F_{\varkappa+2}$ can be obtained from the differential equation satisfied by the hypergeometric function

$$\cos\theta F_{\varkappa+1} = F_\varkappa - \sin^2\theta\frac{A_{\varkappa+1}^\ell}{2(\varkappa+1)}F_{\varkappa+2} \tag{I-10}$$

Using (I-10) in (I-7), we find

$$\Lambda_1 D_\ell^{\varkappa+} = N_{\ell\varkappa}\cos\varkappa\psi\left[2\varkappa(\varkappa-1)\sin^{\varkappa-2}\theta\cos^2\theta F_\varkappa - \varkappa\sin^\varkappa\theta F_\varkappa + 2\varkappa\sin^\varkappa\theta A_\varkappa^\ell F_\varkappa + \left(1-\frac{\varkappa}{\varkappa+1}\right)\sin^{\varkappa+2}\theta A_\varkappa^\ell A_{\varkappa+1}^\ell F_{\varkappa+2}\right] \tag{I-11}$$

Similarly,

$$\Lambda_2 D_\ell^{\varkappa+} = -N_{\ell\varkappa}\sin\varkappa\psi\left[2\varkappa(\varkappa-1)\sin^{\varkappa-2}\theta\cos^2\theta F_\varkappa - \varkappa\sin^\varkappa\theta F_\varkappa + 2\varkappa A_\varkappa^\ell\sin^\varkappa\theta F_\varkappa - \frac{\varkappa}{\varkappa+1}\sin^{\varkappa+2}\theta A_\varkappa^\ell A_{\varkappa+1}^\ell F_{\varkappa+2}\right] \tag{I-12}$$

Multiply (I-12) by $\sin 2\psi$ and (I-11) by $\cos 2\psi$ and subtract to get

$$(\sin 2\psi\Lambda_2 - \cos 2\psi\Lambda_1)D_\ell^{\varkappa+} = -N_{\ell\varkappa}\cos(\varkappa-2)\psi G - \frac{N_{\ell\varkappa}}{N_{\ell\varkappa+2}}A_\varkappa^\ell A_{\varkappa+1}^\ell D_\ell^{(\varkappa+2)+} \tag{I-13}$$

where

$$G = [2\varkappa(\varkappa-1)\sin^{\varkappa-2}\theta\cos^2\theta - \varkappa\sin^\varkappa\theta + 2\varkappa A_\varkappa^\ell\sin^\varkappa\theta]F_\varkappa + \frac{1-\varkappa}{2(1+)\varkappa}\sin^{\varkappa+2}\theta A_\varkappa^\ell A_{\varkappa+1}^\ell F_{\varkappa+2} \tag{I-14}$$

Let $\varkappa\rightarrow\varkappa+1$ in (I-10), then

$$\cos\theta F_{\varkappa+2} = F_{\varkappa+1} - \frac{\sin^2\theta}{2(\varkappa+2)}A_{\varkappa+2}^\ell F_{\varkappa+3}$$

Substituting in the above for $F_{\varkappa+1}$ and $F_{\varkappa+3}$ by using (I-10) we get after some rearrangement

$$\left[\cos^2\theta\frac{(\varkappa+2)}{(\varkappa+1)} + \frac{\varkappa+2}{2(\varkappa+1)^2}\sin^2\theta A_{\varkappa+1}^\ell + \frac{\sin^2\theta A_{\varkappa+2}^\ell}{2(\varkappa+1)}\right]F_{\varkappa+2} - \sin^4\theta\frac{A_{\varkappa+2}^\ell A_{\varkappa+3}^\ell}{4(\varkappa+1)(\varkappa+3)}F_{\varkappa+4} = \frac{\varkappa+2}{\varkappa+1}F_\varkappa$$

Letting $\varkappa \to \varkappa - 2$ multiplying by $2(\varkappa-1)^2 \sin^2\theta$ and rearranging we have for $\varkappa \geq 2$

$$G = 2\varkappa(\varkappa-1)\sin^{\varkappa-2}\theta F_{\varkappa-2} \qquad \text{(I-15)}$$

Also we find directly from (I-11) and (I-12)

$$G = \tfrac{1}{2}\sin^2\theta A_0^{\ell} A_1^{\ell} F_2 \quad \varkappa = 0 \qquad \text{(I-16)}$$

$$G = -(1-2A_{\varkappa}^{\ell})\sin\theta F_1 \quad \varkappa = 1 \qquad \text{(I-17)}$$

Finally then with the substitution of the above in (I-13) we obtain for $\varkappa \geq 2$

$$(\sin 2\psi\Lambda_2 - \cos 2\psi\Lambda_1) D_{\ell}^{\varkappa+} = -2\varkappa(\varkappa-1) B_{\ell\varkappa} D_{\ell}^{(\varkappa-2)+} - \frac{A_{\varkappa}^{\ell} A_{\varkappa+1}^{\ell}}{2B_{\ell,\varkappa+2}} D_{\ell}^{(\varkappa+2)+} \qquad \text{(I-18)}$$

where $B_{\ell\varkappa}$ has already been defined in Eq. (68).

The special cases $\varkappa = 0$, 1 can be determined from (I-13), (I-16), (I-17). With proper normalization

$$(\sin 2\psi\Lambda_2 - \cos 2\psi\Lambda_1) D_{\ell}^{0+} = -\sqrt{2}\frac{A_0^{\ell} A_1^{\ell}}{B_{\ell 2}} D_{\ell}^{2+} \qquad \text{(I-19)}$$

and

$$(\sin 2\psi\Lambda_2 - \cos 2\psi\Lambda_1) D_{\ell}^{1+} = (1-2A_1^{\ell}) D_{\ell}^{1+} - \frac{A_1^{\ell} A_2^{\ell}}{2B_{\ell 3}} D_{\ell}^{3+} \qquad \text{(I-20)}$$

We can combine (I-18, I-19, I-20) to get Eq. (65)a. If $\ell - \varkappa$ is odd, one can show that the + and − superscripts become reversed as is implicit in (65a)

Similarly, we can prove (65b), (66a), (66b).

# APPENDIX II

We can form the raising and lowering operators $M_{\pm}=M_x\pm iM_y$ by using Eqs. (36) and (37). It can be proved easily

$$\frac{i}{\hbar}M+D_l^{m,k}=-\sqrt{(l-m)(l+m+1)}D_l^{m+1,k} \quad \text{(II-1)}$$

$$\frac{i}{\hbar}M-D_l^{m,k}=\sqrt{(l-m)(l-m+1)}D_l^{m-1,k} \quad \text{(II-2)}$$

It is well known that $M_{\pm}$ commutes with $H$, and in particular with the kinetic energy. One can show explicitly that they commute with the relevant part of the kinetic energy given below:

$$[\sin 2\psi\Lambda_2-\cos 2\psi\Lambda_1,\ M_{\pm}]=0 \quad \text{(II-3)}$$

$$[\sin 2\psi\Lambda_1+\cos 2\psi\Lambda_2,\ M_{\pm}]=0 \quad \text{(II-4)}$$

It may be useful to give the following relations

$$[\Lambda_1,\ M_{\pm}]=\pm\frac{2\hbar}{\sin\theta}e^{\pm i\Phi}\Lambda_2 \quad \text{(II-5)}$$

$$[\Lambda_2,\ M_{\pm}]=\mp\frac{2\hbar}{\sin\theta}e^{\pm i\Phi}\Lambda_1 \quad \text{(II-6)}$$

Below we give the results of the raising and lowering operators on the exchange vector spherical harmonics. These results may be derived from Eqs. (II 1), (II 2) and the definition (54).

$$\frac{i}{\hbar}M_+D_l^{(m,\kappa)\pm}=-\sqrt{(l-m)(l+m+1)}D_l^{(m+1,\kappa)\pm} \quad \text{(II-7)}$$

$$\frac{i}{\hbar}M_-D_l^{(m,\kappa)\pm}=\sqrt{(l+m)(l-m+1)}D_l^{(m-1,\kappa)\pm} \quad \text{(II-8)}$$

Operating $M\pm$ on Eqs. (65a), we find

$$\begin{aligned}M_+&(\sin 2\psi\Lambda_2-\cos 2\psi\Lambda_1)D_l^{\kappa+}\\&=-2\kappa(\kappa-1)B_{l\kappa}(1-\delta_{0\kappa})M_+D_l^{(\kappa-2)+}\\&\quad+\delta_{1\kappa}(1-2A_\kappa^l)M_+D_l^{\kappa+}\\&\quad-\frac{-1+3\delta_{0\kappa}}{(1+\delta_{0\kappa}(\sqrt{2}-1))}\frac{A_\kappa^lA_{\kappa+1}^l}{2B_{l,\kappa+2}}\times M_+D_l^{(\kappa+2)+}\end{aligned} \quad \text{(II-9)}$$

Using Eqs. (II 3) and (II 7), we get

$$\begin{aligned}(\sin 2\psi\Lambda_2&-\cos 2\psi\Lambda_1)D_l^{(1,\kappa)+}\\&=-2\kappa(\kappa-1)B_{l\kappa}(1-\delta_{0\kappa})D_l^{(1,\kappa-2)+}\\&\quad+\delta_{1\kappa}(1-2A_\kappa^l)D_l^{(1,\kappa)+}\\&\quad+\frac{(1+3\delta_{0\kappa})}{(1+\delta_{0\kappa}(\sqrt{2}-1))}\frac{A_\kappa^lA_{\kappa+1}^l}{2B_{l,\kappa+2}}D_l^{(1,\kappa+2)+}\end{aligned} \quad \text{(II-10)}$$

With repeated operation of $M_+$, it follows by induction that the above relation is true for any $m$. Similarly we can obtain Eqs. (65b), (66a) and (66b) for any $m$. Thus the radial equations are independent of $m$.

It can also be shown that

$$\psi_{l-m}=(-1)^{\kappa-m}\psi_{lm}^* \quad \text{(II-11)}$$

In order to show this, it is only necessary to state the easily derivable relation

$$[D_l^{(m,\kappa)\pm}]^*=(-1)^{\kappa-m}D_l^{(-m,\kappa)\pm} \quad \text{(II-12)}$$

## REFERENCES


1. E. A. Hylleraas, Z. Physik **48**, 469 (1928); Z. Physik **54**, 347 (1929).
2. G. Breit, Phys. Rev. **35**, 569 (1930).
3. A clear exposition of Breit's work is contained in P. M. Morse and H. Feshbach, *Method of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 1719 et seq.
4. U. Fano, private communication. One of us (A. T.) acknowledges valuable discussions with Dr. Fano in 1959 at the outset of the formulation of these ideas.
5. The D-wave equations in B-H angles have been worked out by H. Feshbach, M.I.T. thesis (1942, unpublished and by one of us, A. Temkin (1959, unpublished). The wave functions for several states have been derived by C. Schwartz, Phys. Rev. 123, 1700 (1961).
6. B. Holmberg, Kungl. Fysiografiska Sallskapets I Lund Förhandlingar **26**, 135 (1956).

7. G. H. DERRICK and J. M. BLATT, Nuclear Phys. **8**, 310 (1958). G. H. DERRICK, Nuclear Phys. **16**, 405 (1960). The characteristic of the three (or more) body problem of finite masses is that six coordinates can be eliminated, the three additional constants of the motion coming from the center of mass. Two papers dealing with the elimination of the six coordinates of rotation and translation from a system of N particles are:
J. O. HIRSCHFELDER and E. WIGNER, Proc. Nat. Acad. Sci. **21**, 113 (1935); C. F. CURTISS, J. O. HIRSCHFELDER, and F. T. ADLER, J. Chem. Phys. **18**, 1638 (1950).
8. E. WIGNER, *Group Theory*, (Academic Press, New York and London, 1959), p. 216. cf. also, A. R. EDMONDS, *Angular Momentum in Quantum Mechanics*, (Princeton University Press, Princeton, 1957), Chapter 4.
9. W. MAGNUS and I. OBERHETTINGER, *Formulas and Theorems for the Functions of Mathematical Physics*. (Chelsea Publishing Co., New York, 1949).
10. L. PAULING and E. B. WILSON, Introduction to Quantum Mechanics, (McGraw-Hill Book Co., Inc., New York and London, 1935), p. 280.
11. A. TEMKIN, Phys. Rev. 126, 130 (1962), P. LUKE, R. MEYEROTT, W. CLENDENIN, Phys. Rev. **86**, 401 (1952).
12. A. TEMKIN and E. SULLIVAN, Phys. Rev. 129, 1250 (1963).
13. A. W. WEISS, Phys. Rev. **122**, 1826 (1961).
14. C. SCHWARTZ, Phys. Rev. **126**, 1015 (1962).
15. A. W. WEISS and J. B. MARTIN, Phys. Rev. **132**, 2119 (1963). For a somewhat different finding from those of reference 13 and 14 and the preceding paper of this reference, cf. W. LAKIN, A.E.C. Report NYO-10, 430 (1963, unpublished).
16. A. TEMKIN, Phys. Rev. Letters, **6**, 354 (1961).
17. C. L. PERKERIS, Phys. Rev. **112**, 1649 (1958).
18. C. L. PERKERIS, Phys. Rev. **115**, 1216 (1959). C. SCHWARTZ, Phys. Rev. **128**, 1146 (1962).
19. A. TEMKIN, Proc. Phys. Soc. (London) **80**, 1277 (1962).
20. A. TEMKIN, J. Chem. Phys. 39, 161 (1963).
21. A. TEMKIN and A. K. BHATIA (to be published).

# SPACE: HIGHLIGHTS OF RECENT RESEARCH*

ROBERT JASTROW AND A. G. W. CAMERON

Space science is the collection of scientific problems to which space vehicles can make some specific contributions not achievable by ground-based experiments. At present this field includes broad segments of the traditional disciplines of the earth sciences, physics, and astronomy. In future years the biological sciences will join this group in an important role, as our explorations of the moon and planets provide us with opportunities for studying the conditions under which physical life may have developed. This article reviews some highlights of recent space research in the physical sciences.

## GEODESY

Important results have been achieved in determining the internal structure of our own planet with the aid of near earth satellites. A satellite's orbit is determined by the distribution of mass within the earth. If the earth were a perfect sphere, under the attraction of the mass point at the earth's center of gravity the satellite would move in an ellipse whose plane would have a constant orientation in space.

Actually, the plane of a satellite's orbit rotates slowly in space, due to the additional force of attraction exerted by the equatorial bulge. Studies of the orbital rotation rates of a number of satellites have yielded a very precise value for the height of the equatorial bulge. These indicate a discrepancy between the observed value for the flattening and the value expected on the basis of an assumption of hydrostatic equilibrium. It has been suggested by Munk and MacDonald that these results imply that the interior of the earth is not in hydrostatic equilibrium but has a mechanical strength sufficient to maintain the earth's shape in spite of the stresses at the base of the mantle.

There are other departures of the geoid from the shape expected in the case of hydrostatic equilibrium, in addition to the discrepancy in the flattening. These departures are of very great significance because they represent variations in gravity which depend on the entire distribution of mass within the planet, which are more significant for the gross structure of the planet than the simple topographical variations—for example, mountains—which represent the distribution of the mass at the surface alone.

Detailed analysis of these gravitational variations yields a figure of the earth having a positive anomaly, or a lump, in the region of the western Pacific, near Indonesia and the Philippines, a large depression or negative anomaly in the Indian Ocean, and a negative anomaly in the Antarctic (Fig. 1).

Although these depressions and elevations are relatively minute, they are exceedingly significant because they represent variations in the force of gravity, or in the amount of matter per square centimeter, in the regions in question. For example, the depression in the Indian Ocean is only 60 meters deep, but it signifies that the force of gravity there is, relatively, so weak that the waters of the sea are not drawn together to the depth that one would expect if the whole earth were subject to a uniform gravitational force.

These anomalies are correlated with the rate at which heat flows through the body of the earth to the surface. The correlation is such that, where the geoid is anomalously high, the heat flow is anomalously low. On the average, the flow of heat outward through the crust of the earth is 60 erg $cm^{-2}$ $sec^{-1}$. In the depression of the geoid near India, the flow of heat is substantially higher, 80 erg $cm^{-2}$ $sec^{-1}$. At the elevation of the geoid in the western Pacific, the flow of heat is substantially lower, about 40 erg $cm^{-2}$ $sec^{-1}$.

We expect this kind of correlation if there is a

*Published in *Science*, 145 (3637): 1129–1139, September 11, 1964.

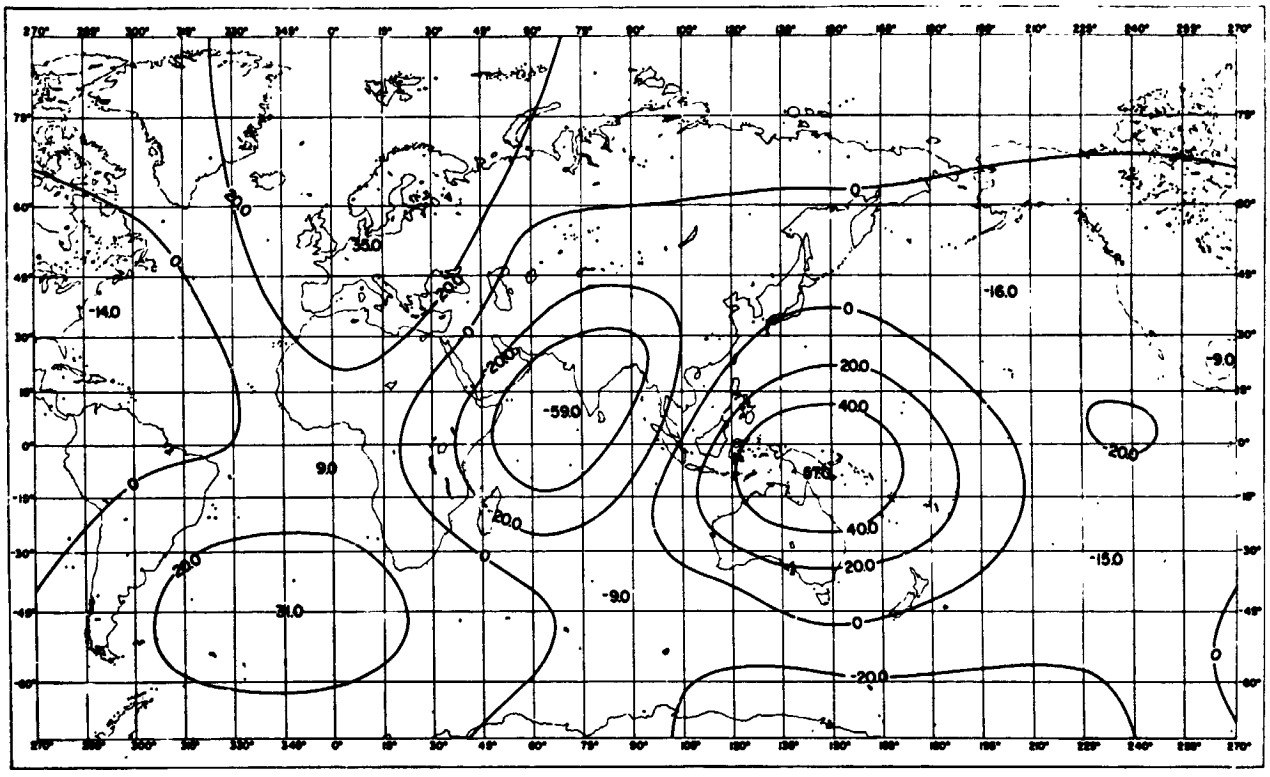


FIGURE 1.—Geoid heights (in meters) relative to an ellipsoid with a flattening of 1/298.3. The major features include a negative anomaly in the Indian Ocean and a positive anomaly centered near Indonesia and the Philippines. [From G. J. F. MacDonald, *Science* 143, 921 (1964)].

mass transport, or convection of matter, from the deep interior of the earth to the surface in these regions. If there were an upward motion through the interior of the earth which carried relatively warm material from below to the surface, this upward-moving column would have a lower density than its surroundings, and therefore the mass per square centimeter in the column, and the gravitational force on the surface of the earth about it, would be lower than the average. At the same time, the heat which the warm column transports upward would add to the normal release of radioactive heat throughout the mantle and crust, so that above the same upward-moving column there would be an exceptionally high rate of heat flow through the surface. The converse would hold for a descending column, which would carry a relatively dense, and therefore relatively cold, material from the surface layer to the interior of the earth. Above the cold and dense column the gravitational force would be relatively great, and a bump would appear in the sea level there. This may be the cause of the elevation in the western Pacific.

## METEOROLOGY

In geocentric order, the next major area of investigation in space science concerns the atmosphere and the control exerted over it by the sun. This field of research includes questions related to the circulation of the winds in the lower atmosphere and to the vertical structure of the atmosphere at higher altitudes.

Eight Tiros satellites have been launched in the past 4 years, all carrying vidicon cameras for the global study of the cloud cover; Tiros II, III, IV, and VII carried, in addition, a set of infrared detectors for measuring the intensity of infrared radiation emitted from the earth-atmosphere system.

The cloud-cover photographs have already yielded results that are of great interest when correlated with ground observations, and they indicate that use of satellites may lead to a substantial improvement in weather forecasting by providing global and nearly continuous coverage of regions of weather activity. The matter of global coverage is critically important, because the success of weather forecasting has been found

to increase rapidly with the size of the region covered by the observations; yet at present large parts of the globe are very poorly covered. These are regions in which weather activity can develop and grow without detection before moving out into the inhabited areas. They include the polar regions, the major deserts, and the southern oceans. Satellite coverage will greatly strengthen the hand of the meteorologist by filling in these blank portions of the global weather map, and it may be expected to have important consequences for the economies of this country and the world.

The cloud photographs may also be important for the basic objectives of long-range forecasting and the understanding of the causes of weather activity, because clouds have a strong influence on both the amount of solar energy admitted to the earth-atmosphere system and the amount of energy returned to space from the surface.

The energy balance of the earth-atmosphere system is the difference between the incoming solar radiation, mostly in the visible region of the spectrum, and the outgoing terrestrial radiation, in the infrared. The latitudinal variation of the energy balance shows an excess of incoming solar radiation over outgoing radiation near the equator, and a deficiency at the poles. It is this variation of the energy balance with latitude that drives the atmospheric heat engine.

That part of the visible radiation from the sun which is not reflected by clouds or scattered in the atmosphere reaches the surface of the earth and is absorbed, heating the ground to a temperature in the neighborhood of 235°K. For a glowing body at a temperature of 235°K, most of the energy is radiated at wavelengths in the far infrared. This infrared radiation is strongly absorbed by several constituents of the atmosphere, including water, carbon dioxide, and ozone. The absorption of infrared radiation from the ground by these molecules heats the lower atmosphere, which reradiates the absorbed energy, partly upward to outer space and partly downward, providing additional heating of the surface.

The additional heating of the surface by the return of infrared radiation from the atmosphere is analogous to the action of the glass panes of a greenhouse, and is called the "greenhouse effect." It is sufficient to raise the temperature of the surface of the earth by about 55°K, so that the average temperature of the surface of our planet becomes 290°K.

Clouds are strong absorbers of infrared radiation and influence the infrared transmission and the local greenhouse effect. (This is in addition to the primary effect of clouds as reflectors of incident visible solar radiation.)

Thus, the cloud cover has an important effect on the deposition of energy in the atmosphere because it influences both the inflow and the outflow of energy through the atmosphere. Thus far the characteristics of clouds—amount, types, and approximate heights—have all been measured by ground-based observers. Satellite observation by television cameras introduces the possibility of obtaining extensive data on global cloud cover in relatively short periods.

A. Arking, of the Goddard Institute for Space Studies and New York University, has recently carried out the first statistical analysis of Tiros cloud-cover data, using a digitizing technique applied to Tiros III photographs. The Tiros III results are compared in Fig. 2 with a climatological mean cloudiness for the Northern Hemisphere that was compiled by K. Telegadas and J. London, then at New York University, from ground observations extending over a 50-year period. The Southern Hemisphere data of Fig. 2 were compiled by H. Landsberg of the U.S. Weather Bureau.

The results in Fig. 2 show that the cloud cover in middle latitudes is the same in the Northern and Southern hemispheres. However, in tropical latitudes there is an asymmetry, with a local maximum of the cloud cover in the tropics centered at 10° N. This is the average position of the "thermal equator" during the period 12 July to 30 September.

These results are preliminary, but the approach seems promising, and it is hoped that improved techniques will make it possible to detect fine-scale temporal and geographical variations in the distribution of the energy balance.

## UPPER ATMOSPHERE

The physical processes which control the upper atmosphere are determined largely by the absorption of solar ultraviolet radiation by the atoms and molecules existing at great heights. Although the ultraviolet component of the solar radiation

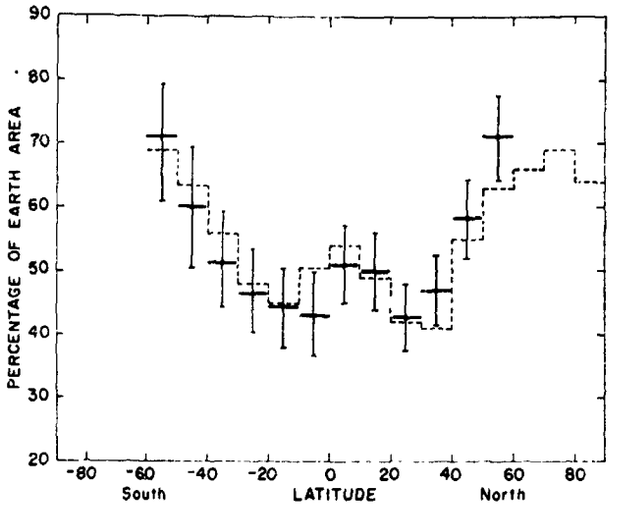


FIGURE 2.—The latitudinal distribution of cloud cover. (Solid horizontal bars.) Results derived from Tiros III photographs taken from 12 July to 30 September 1961; (vertical lines) estimated uncertainty in the Tiros derived data; (dashed histogram) climatological mean cloud cover derived from ground observations, as compiled by Telegadas and London for the Northern Hemisphere and by Landsberg for the Southern Hemisphere. The broad features of the latitudinal distribution are consistent with the known pattern of the general circulation. The air rising at the thermal equator produces a relative maximum in the cloud cover, while, on the average, there is downward motion of cool, dry air at latitudes 30° north and south of the thermal equator, which explains the realtive minimum of cloudiness. (From A. Arking, *Science* 143, 569 1964)

is only a small fraction of the total flux of solar energy, the absorption cross sections in the far ultraviolet are so large that radiation at these wavelengths has been effectively removed from the incident spectrum by the time the incident flux has penetrated to a height of 100 kilometers. The ultraviolet radiation is the principal source of heating of the thin upper air and the major determining factor in its structure.

At lower altitudes the air is composed of oxygen and nitrogen, and we can measure the proportions of these rather accurately. At the highest altitudes these gases have partially settled out of the air through diffusion. The lighter gases dominate the composition of the air at sufficiently high altitudes. Of these gases hydrogen is the lightest, and for this reason it was once believed to be the dominant constituent of the air above the oxygen-nitrogen layer. The emergence of the hydrogen atmosphere was thought to come at an altitude of about 1200 kilometers. However, in July 1961, Marcel Nicolet of Belgium suggested, on the basis of an initial examination of the density data of Echo I, that between the oxygen-nitrogen atmosphere and the hydrogen atmosphere there lies a layer of helium. The helium layer was discovered experimentally a short time later by R. Bourdeau, of the NASA Goddard Space Flight Center, and the finding has since been confirmed in other experiments (Fig. 3).

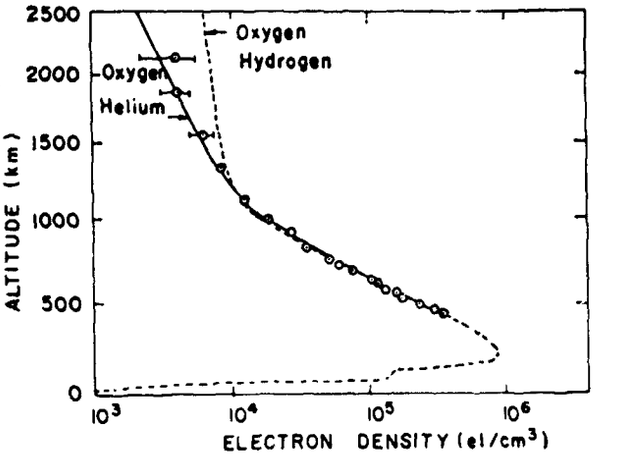


FIGURE 3.—Electron density profile measured by a Scout rocket, and providing evidence for a helium layer in the upper atmosphere. (Dotted line) Density distribution for a scale height derived from an oxygen-hydrogen mixture; (solid line) calculated density distribution for an oxygen-hydrogen mixture. [From S. J. Bauer and J. E. Jackson, J. Geophys. Res. 67, 1676 (1962)]

Our knowledge of properties of the atmosphere at altitudes above 200 kilometers is mainly derived from measurements of atmospheric drag on satellites. The period of revolution of a satellite decreases steadily at a rate proportional to the drag force exerted by the atmosphere, which is, in turn, proportional to air density; a measurement of the rate of change of period therefore gives the value of the air density suitably averaged around the orbit.

Data on satellite drag have been a very valuable source of information on atmospheric properties. L. G. Jacchia of the Smithsonian Astrophysical Observatory was the first to discover, by careful analysis of time variations in the drag, that the upper atmosphere is extremely responsive to solar control, deviating in density from the mean by as

much as a factor of 100 and deviating in temperature by hundreds of degrees at times of solar activity.

The significance of the correlation between solar activity and the properties of the earth's upper atmosphere can be described as follows. The surface of the sun is the scene of great activity, especially during the maximum of the sunspot cycle, when it is marked by sunspots and by hot, dense regions with temperatures of some millions of degrees, which are located in the solar corona above the sunspot areas. When such an active region faces toward the earth in the course of the sun's rotation, extreme ultraviolet radiation emitted from these active regions is absorbed in the upper atmosphere. The precise correlation between solar activity and density was discovered by Jacchia and by W. Priester of Bonn University Observatory and the Goddard Institute for Space Studies. Their results suggest that the amount of energy transferred to the earth is sufficient to heat the atmosphere appreciably, causing an upward expansion and a large increase in the density of the exceedingly thin air at high altitudes. This discovery provided the first direct evidence of the influence of solar surface activity on atmospheric properties.

The continuing analysis of the correlation has given us a rather full picture of solar control over the upper atmosphere. It indicates that the atmosphere is appreciably heated by the ultraviolet radiation emitted at times of general solar surface activity, and is further heated by interaction of the earth with the clouds of solar particles which are emitted from the sun following solar surface eruptions. The arrival of the clouds of particles at the earth is signified by the onset of geomagnetic disturbances or "magnetic storms." It is found that increases in the temperature of the atmosphere occur shortly after the commencement of the magnetic storms. Thus, it appears that both ultraviolet radiation and corpuscular streams are sources of energy for the upper atmosphere (Figs. 4 and 5). The question of the energy sources for the upper atmosphere is the most important single problem for upper-atmosphere physics at this time. The continuing investigation of this matter and, in particular, of the roles played by particle and radiation sources, respectively, will be one of the main areas of ex-

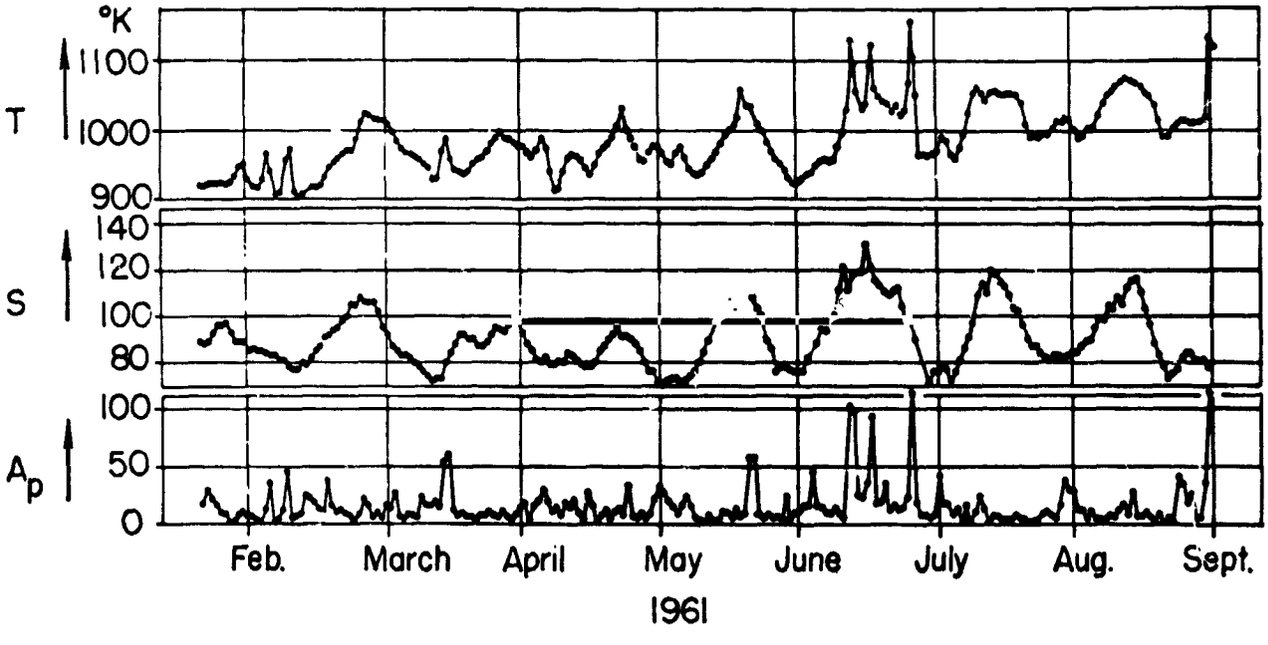


FIGURE 4.—Temperature variation correlated with solar radio emission and geomagnetic indices for the interval from February through September 1961. (Upper curve) Exospheric densities and temperatures, derived from Explorer IX drag measurements by L. G. Jacchia and J. Slowey (1962); (middle curve) flux of the solar radiation at 20-centimeter wavelengths, an indicator of solar activity measured at the Heinrick Hertz Institut in Berlin; (lower curve) geomagnetic activity index ($A_p$). [The Explorer IX data are adapted from L. G. Jacchia and J. Slowey. *Smithsonian Astronomical Observatory Spec. Rept. No.* 84 (1962); reproduced courtesy of W. Priester.]

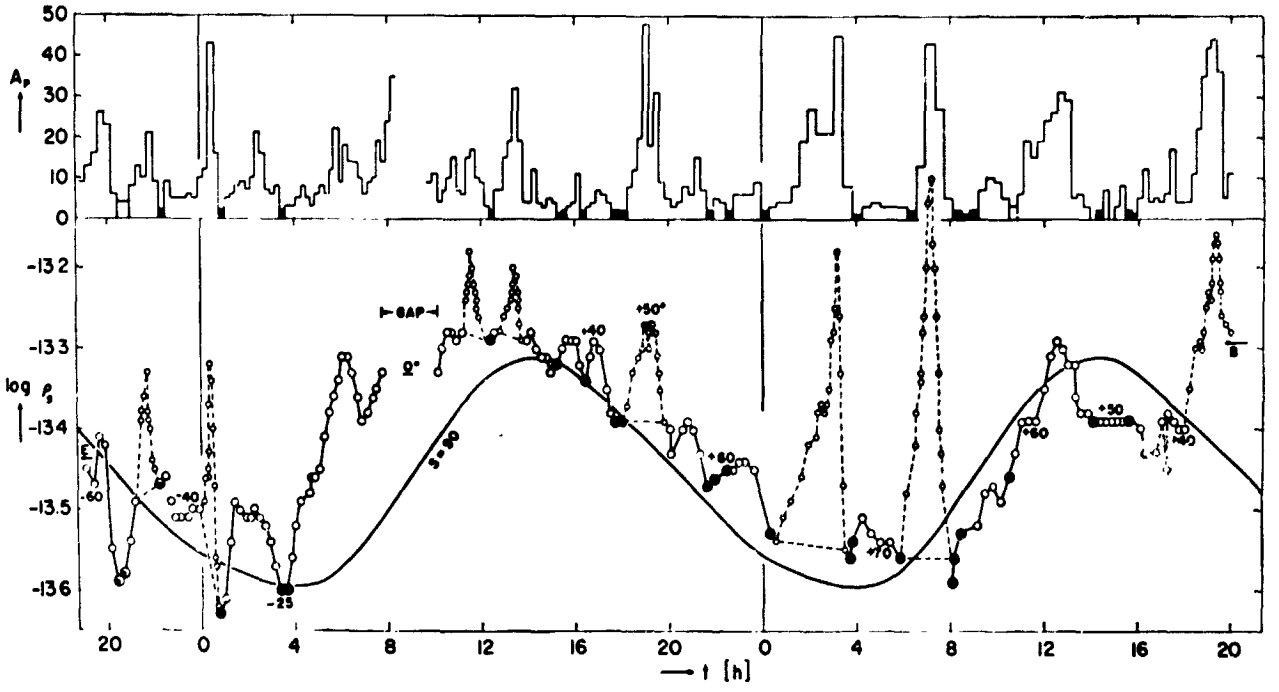


FIGURE 5.—Density variations at 250 kilometers above sea level as a function of local time $t$, determined by L. G. Jacchia and J. Slowey (1963) of the satellite Injun III from 15 December 1962 through 29 June 1963. During this time the geographic latitude of the perigee cover the range from +70° to −60°, as indicated by the numbers on the density curve. (Solid curve) The Harris-Priester model for the proper level of solar activity; (histogram at top) daily geomagnetic indices $A_p$; (open circles) densities during magnetic storms; (solid circles) densities during magnetically quiet days ($A_p<2$). The response of the atmosphere to solar activity, as indicated by violent solar storms, is much greater within the auroral zone than outside it. [Adapted from L. G. Jacchia and J. Slowey, *J. Geophys. Res.* 69, 905 (1964)].

perimental and theoretical effort in the next several years.

## MAGNETOSPHERE

The evidence cited suggests that corpuscular streams from the sun transfer appreciable amounts of energy to the atmosphere. How does the transfer of energy in the atmosphere occur?

The general answer seems to be connected with the properties of the outermost layer of the atmosphere. The density of the upper air merges into the density of the interplanetary gas at an altitude of about 100,000 kilometers, marking the boundary of the atmosphere. Early in 1958, however, J. A. Van Allen of the State University of Iowa discovered, by analyzing Geiger-counter data from Explorer I, that there was an additional layer of energetic charged particles in the upper atmosphere. These charged particles are trapped in the atmosphere by the earth's magnetic field, and the atmospheric layer which they constitute is therefore called the magnetosphere.

During the last few years three important developments have substantially changed our earlier impressions about the character of the magnetically trapped particles and their geophysical effects.

First, B. O'Brien, also of the State University of Iowa, using measurements from the Injun I satellite, discovered that the flux of charged particles coming down from the trapped-particle region was so large that, if this flux consisted of previously trapped particles which had just been dislodged by solar disturbances, it would drain the whole magnetosphere in about an hour. He also found that when a solar disturbance occurred, both the flux of untrapped descending particles and the number of trapped particles increased. Thus he concluded that the leakage of trapped particles from the Van Allen belts cannot be the principal source of the electrons which pass down through the atmosphere. He decided that, while a few charged particles are trapped during or after a solar disturbance, most pass directly into the atmosphere without spending an appreciable amount of time in the trapped-particle region.

Apparently, the charged particles which are observed in auroral displays and other atmospheric phenomena are those which come directly down the lines of force into the atmosphere.

Second, a large population of low-energy protons, having a range from 100,000 to several million electron volts, was discovered by A. H. Davis and J. M. Williamson of the Goddard Space Flight Center. The concentration of these protons reaches a maximum at a distance of 3.5 earth radii. At that distance their density is about one per cubic centimeter. This value for density of the trapped protons has interesting implications. As a result of the magnetic field gradient and curvature effects, the trapped protons drift westward in the magnetic field, with an associated electric current that produces magnetic effects. These effects have been calculated by S. Akasofu of the University of Alaska, S. Chapman of the universities of Colorado and Alaska, and (in unpublished work) R. A. Hoffman of the Goddard Space Flight Center. They find that the changes in the intensities of these trapped protons produce magnetic perturbations large enough to explain most magnetic storms observed on the earth, and also the very large perturbations of the geomagnetic field in space in the neighborhood of the proton belt. The relation between the trapped-proton drift current and the geomagnetic storms was suggested by S. F. Singer of the University of Maryland in 1956.

The third development was the discovery of a substantial flux of electrons with very high energies (in the neighborhood of 1 million volts) at a distance of 3 or 4 earth radii, presumably produced by beta decay of albedo neutrons resulting from cosmic ray interactions in the atmosphere. These electrons penetrate the Geiger counters with high efficiency, and when allowance is made for their presence, the estimate of the total flux of electrons is reduced from the earlier estimated value of $10^{10}$ cm$^{-2}$ sec$^{-1}$ srad$^{-1}$ to the currently accepted value of $10^{8}$ cm$^{-2}$ sec$^{-1}$ srad$^{-1}$.

## THE MAGNETOPAUSE

The connection between the magnetosphere and the transfer of corpuscular energy to the atmosphere is probably to be found in the properties of the atmosphere near the magnetopause, the boundary separating the interplanetary medium from the region around the earth, in which the geomagnetic field is dominant. The sharply defined surface of the magnetopause marks the termination of both the trapped-particle region and the geomagnetic field. Satellite measurements of the geomagnetic field by L. J. Cahill of the National Aeronautics and Space Administration and of the University of New Hampshire show that the magnetopause has a thickness on the order of 100 kilometers and occurs at a distance of 8 to 10 earth radii on the sunlit side of the earth.

The sharpness of the magnetospheric boundary is illustrated by Figs. 6 and 7. Figure 6 represents the magnetic-field measurements obtained by Cahill and P. G. Amazeen of the University of New Hampshire, using a three-component magnetometer flown on Explorer XII. At the magnetopause a sudden drop occurs and outside the magnetosphere the magnetic field is highly variable in magnitude and direction. Figure 7 shows the counting rates of charged particle detectors flown on Explorer XIV by Van Allen, L. A. Frank, and E. Macagno of the State University of Iowa.

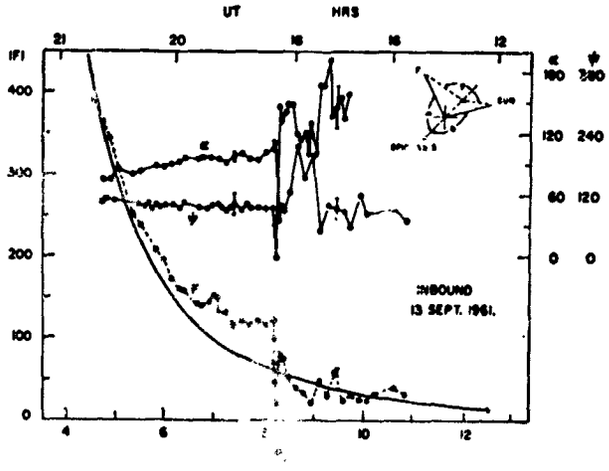


FIGURE 6.—Explorer XII measurements showing the abrupt termination of the geomagnetic field at the magnetopause. $F$ is the magnitude of the magnetic field and $\alpha$ and $\beta$ refer to its direction. Within the magnetosphere the field is closely described by the high-altitude extrapolation of the earth's approximately dipole field, shown as the solid curve. Outside the magnetopause, which occurred at 8.2 earth radii ($R_E$) during this flight, the field varies in magnitude and direction. [From L. J. Cahill and P. G. Amazeen, *J. Geophys. Res.* 68, 1841 (1963)].

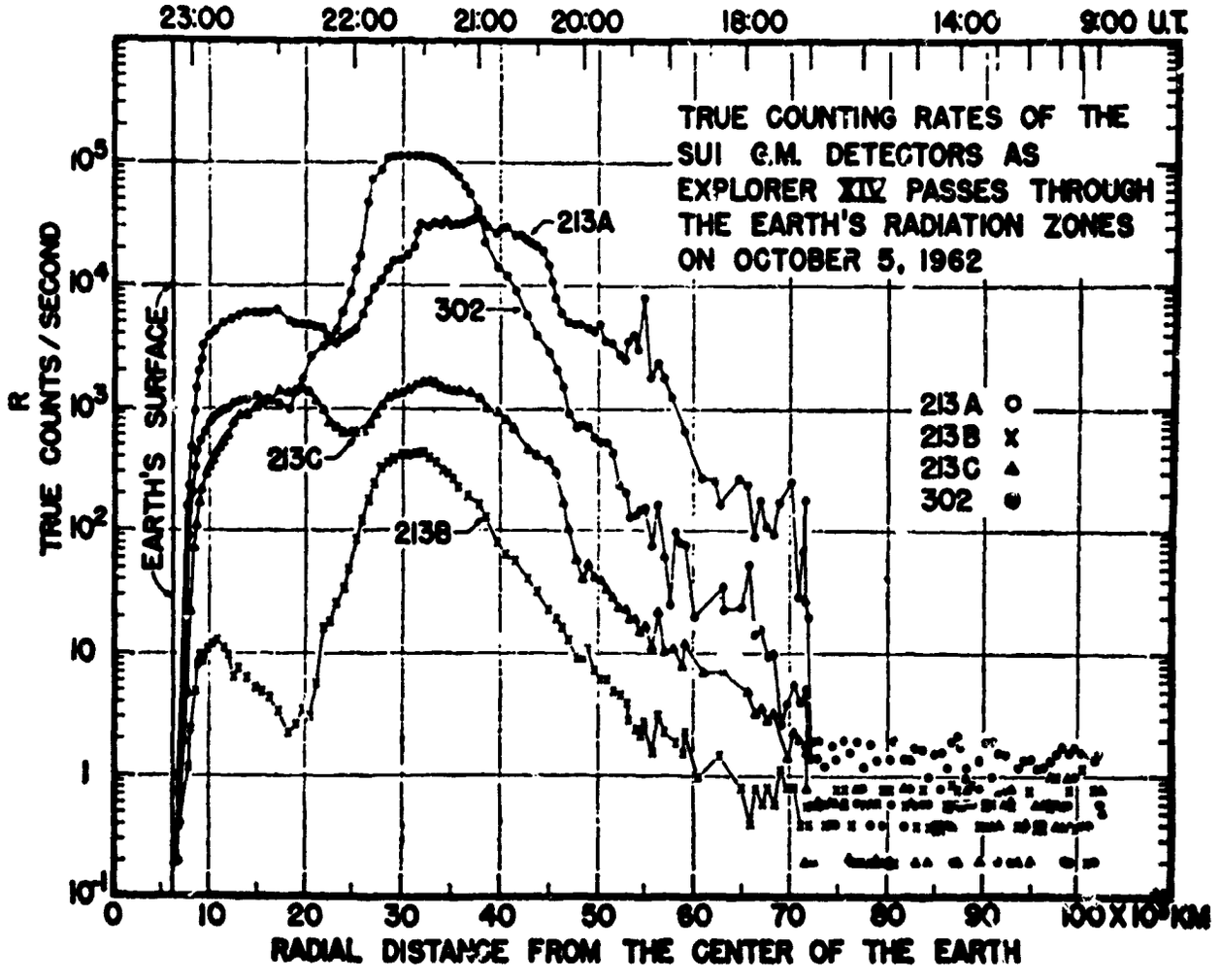


FIGURE 7.—Explorer XIV data showing the abrupt termination of the magnetically trapped particle belts at the magnetopause. The detector labeled 213A accepts the particles of lowest energy; it records principally 50-kev electrons. Its counting rate drops sharply at a radius of 72,000 kilometers, or 12.5 $R_E$, which corresponds to the location of the magnetopause during this flight. [From L. A. Frank, J. A. Van Allen, E. Macagno, *J. Geophys. Res.* 68, 3545 (1963).]

The detector which accepts the particles of lowest energy is labeled 213A, and its counting rate reflects principally the flux of 50-kev electrons. At the magnetopause the counting rate of this detector drops, by a factor of approximately 20, to a value which is approximately independent of altitude and is produced by the cosmic ray background in space.

Within the magnetopause there are no substantial fluxes of energetic particles other than those of the magnetically trapped particles illustrated in Fig. 8. Detectors flown on Mariner II indicate that the sun is the source of a particle stream which flows through interplanetary space continually, although with variable velocity and intensity. The Mariner II detectors, and also the plasma probe flown on Explorer X by B. Rossi, H. S. Bridge, A. J. Lazarus, A. Bonetti, and F. Scherb of the Massachusetts Institute of Technology, have shown that these particles move radially outward from the sun at velocities varying from 300 to 600 kilometers per second and at an average flux of $10^8$ $sec^{-1}$ outside the magnetopause. The interplanetary stream of solar particles, called the solar wind, cannot penetrate the magnetic field of the earth but divides and flows around it as the waters of a stream divide around a boulder (Fig. 8). The closest distance of approach of the solar wind to the earth is about 10 earth radii.

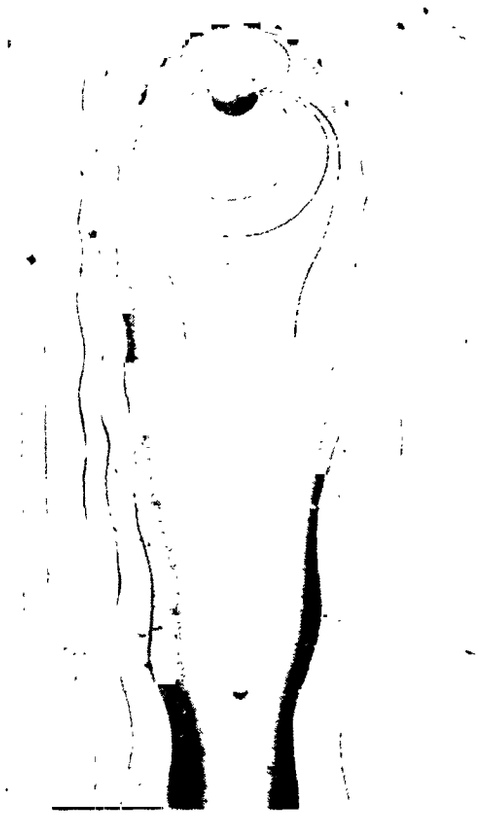

FIGURE 8.—The geomagnetic cavity in the solar wind.

The shadow or cavity carved by the magnetic field of the earth in the solar wind should, in principle, extend back indefinitely far into the solar system behind the earth. However, because the particles of the stream have appreciable transverse velocities associated with their thermal motions, we expect these particles to diffuse together eventually in the shadow of the earth. The ratio of mean transverse to radial velocities is about ¼, hence we expect the geomagnetic cavity to be filled in at a distance of 4 times the diameter of the cavity, or, roughly, the distance of the moon from the earth, as suggested in Fig. 8.

The amount of energy transferred from the solar wind to the atmosphere within the cavity is difficult to estimate. The variable magnetic fields in the solar plasma are believed to glue the particles together and give their motion the properties of fluid flow, in spite of the low density; thus turbulence would be expected at the region of impact of the solar wind on the magnetopause. The bufleting of the magnetospheric boundary associated with this turbulent impact may generate disturbances in the field just within the magnetopause. These disturbances propagate hydromagnetically down or across the field lines into the atmosphere, where they transfer energy which may appear as atmospheric heating, ionization, auroral disturbances, and magnetic storms —that is, the whole complex of disturbances produced in the atmosphere at high geomagnetic latitudes at times of solar activity.

The interaction of the solar plasma with the magnetosphere also leads to the expectation that a "shock wave" will be formed some distance beyond the actual magnetopause. This expectation arises from the fact that the flow of the plasma is supersonic and must become subsonic in the vicinity of the earth. Such a transition would set up a shock wave which would stand off some distance from the magnetopause and would have a thickness determined by the ability of the magnetic field to change the bulk motion of the plasma particles. Evidence of the perturbed magnetic fields corresponding to the shock wave and the intervening transition region has been detected by N. F. Ness of the Goddard Space Center with magnetometers flown on the Interplanetary Monitoring Platform (IMP) spacecraft launched on 27 November 1963.

## ATMOSPHERE OF VENUS

Venus is the third brightest object in the sky, the planet nearest the earth, and the planet most closely resembling the earth in size and mass. It has been studied with the telescope since Galileo's time, and yet it remains an enigma, because its surface is permanently shrouded by a layer of clouds. In this state of ignorance, hope has flourished that Venus offers a hospital environment for the development of advanced forms of life.

Some information regarding the surface of the planet has been obtained in recent years through study of the microwave radiation emitted in the longwave region of its thermal spectrum. This radiation, with wavelengths in the region of 1 to 10 centimeters, penetrates the clouds without significant attenuation; its intensity is proportional to the temperature of the emitting surface.

The first attempts to measure the microwave radiation from Venus were made in 1956 with the Naval Research Laboratory radio telescope. The

temperature inferred from the measured radiation intensity was approximately 600°K (700°F)—certainly too high to permit any terrestrial forms of life. Repeated measurements have confirmed the Naval Research Laboratory results and have a revision of our ideas regarding the surface and lower atmosphere of Venus.

It is difficult to understand why the temperature of Venus should be so much higher than that of the earth. The path to an explanation would seem to lie in the assumption of an extremely dense atmosphere which absorbs strongly in the infrared region of the spectrum but is transparent with respect to visible radiation. The part of the incident sunlight which is not reflected back by the clouds will therefore penetrate through the atmosphere and heat the surface of the planet. But when the surface layers reradiate this energy at infrared wavelengths, the radiation is absorbed by the atmosphere and returned in large measure to the surface, thus giving an additional flux of energy into the ground and raising its temperature. This atmospheric phenomenon is the greenhouse effect described earlier.

The clouds of Venus reflect three-quarters of the incident sunlight; the remaining quarter of the incident radiation would bring the surface of the planet to a temperature of 235°K if there were no atmospheric greenhouse effect. If the greenhouse effect is to raise the ground temperature to 600°K, the optical thickness of the atmosphere must be 50 mean free paths throughout the far-infrared region. In an atmosphere with this degree of opacity, only one photon in $10^{21}$ escapes directly, without absorption. This condition is so severe that alternative suggestions have been made, among them the hypothesis that the apparently high radio temperature of Venus is the result of microwave emission from its ionosphere rather than from the surface.

The Mariner II Venus flyby launched on 27 August 1962 included experiments designed to test this hypothesis. This spacecraft passed Venus at a distance of 33,440 kilometers (20,900 miles) on 14 December 1962, and made crucial measurements of the temperature across the disc. The spacecraft was equipped with two sets of radiation detectors, one in the infrared and one in the microwave region. Measurements of the radiation emitted by the planet in the microwave region included measurements of radiation of 19-millimeter wavelength, which passes through the atmosphere with little attenuation and hence provides a measure of the temperature at the ground, provided there is no additional emission from the ionosphere.

A modest degree of atmospheric attenuation is, however, to be expected, and in the scan of Mariner II across the disc of the planet this slight degree of attenuation should show up as a lower intensity of measured radiation at the edge or limb of Venus, where the thickness of the intervening atmosphere is greater. However, if the high microwave intensities and apparent temperatures result from emission by electrons in the atmosphere of Venus, then the readings at the limb should indicate an enhancement or brightening because of the greater thickness of the ionosphere in the line of sight.

The Mariner II results showed a conclusive darkening of the limb of Venus at 19 millimeter wavelength, thus eliminating the possibility of ionospheric emission and confirming the supposition that the measured radio temperature of 600°K is associated with the surface of the planet.

## EXPLORATION OF THE MOON

The moon is a uniquely important body in the study of the history of the solar system because its surface has preserved the record of its history remarkably well. The moon has a negligible atmosphere and no oceans. It is, therefore, unchanged by the processes of erosion which erase the history of the earth's surface in a relatively short time—between 10 and 30 million years.

This is evidenced, in part, by the tens of thousands of craters on the lunar surface, produced by the impact of meteorites which, presumably, have been colliding with the moon since its formation. This is perhaps the only physical record which we have of events in the development of the solar system going back to that early time.

Because of the antiquity of the moon's surface, there is another remarkable record preserved—a layer of cosmic dust which is believed to have rained on it from the solar system since its formation. This dust may be as much as 30 centimeters or more in depth and may contain organic molecules and the precursors of life on earth.

The most important measurements of lunar properties from spacecraft have resulted from flights of the Russian Lunik II and Lunik III. From the Lunik II magnetometer data Soviet scientists concluded that an upper limit of approxmately 100 gammas could be placed on the moon's magnetic field. In future flights, refinement of this limiting value for the moon's magnetic field may provide information on the presence or absence of a liquid core within that body (on the earth the magnetic field is supposed to be associated with currents in the liquid core of the planet). This in turn could have a bearing on our understanding of the formation of the moon and of similar bodies in the solar system.

Lunik III has provided us with the first pictures of the remote side of the moon. In spite of some blurring, the photographs are of great interest, for it is possible to distinguish a large number of features resembling the craters and maria on the front face. Perhaps the most interesting feature is the Soviet Mountain Range, a chain extending across the center of the moon's hidden face. It resembles the great ranges on the earth and is unlike the mountain formations characteristic of the moon's front face, which seem to be circular crater walls and deposits of debris formed by the impact of large meteorites on the lunar surface.

According to our present ideas, terrestrial mountains result from the combined effects of erosion and wrinkling of the earth's crust, but such mountain-building forces have been much less effective on the moon. In the photograph, the markings designated the Soviet Mountain Range could have resulted from the running together of several obscured but independent markings. However, if these features continue to appear as a single range in later, more detailed pictures, we may have to revise our theories of lunar structure.

## SOLAR PHYSICS

One of the most interesting questions in solar physics concerns the manner in which energy is transported above the surface of the sun to heat the chromosphere and corona.

We know that near the center of the sun, where the temperature is approximately 15 million degrees Kelvin, hydrogen is converted into helium by a variety of nuclear reactions. We also know that the sun is a self-adjusting system which expands or contracts in order to maintain a precise balance between the energy generation at the center and the energy emission from the surface.

All regular mechanisms of energy transport can carry heat only from a region of high temperature to a region of low temperature. Therefore, in order for the heat generated by nuclear reactions to be carried away from the center of the sun, the temperature must fall continuously from the center to the edge. This is in fact the case; the temperature falls from 15 million degrees at the center of the sun to 5800 degrees at the visible edge.

However, above the visible edge, which is called the photosphere, there lies a relatively tenuous region of gas which constitutes the atmosphere of the sun. This region is divided into the chromosphere and, above that, the corona.

The puzzling fact is that the temperature of the sun *rises* again above the photosphere, reaching a value of 1.5 to 2 million degrees in the corona. One of the burning questions of solar physics is what constitutes the source of the energy which produces the very high temperatures in the solar corona. Also, what is the mechanism of energy transport by which energy can be carried without appreciable losses through the dense gases of the photosphere and yet undergo great losses in the tenuous regions of the corona?

A current belief is that a wave motion—either a sound wave, a hydromagnetic wave, or a gravity wave—carries energy upward from the photosphere and deposits it in the corona. When a sound wave propagates into a region of decreasing density, its amplitude increases and it steepens into a shock wave. This is a mechanism in which considerable dissipation of energy takes place. It appears that hydromagnetic waves are rapidly damped out below the photosphere, but if they could be generated in the region of the chromosphere, then they would tend not to be dissipated until the waves had reached the corona. Magnetic disturbances above the photosphere may be particularly effective in generating these waves. Gravity waves consist of a kind of rolling motion similar to the waves on the surface of the ocean. These may, like sound waves, be generated by the motions of convecting material in the transition layer; they would have a vertical component of

propagation and would be dissipated in the corona.

It may be that all three of these mechanisms are operating in the heating of the chromosphere and corona. If this is the case, there may be a steady heating of the corona upon which is superimposed a localized heating associated with magnetic activity. Thus, the heating of the corona is expected to depend upon the magnetic structure in the outer layers of the sun. This dependence is observed in many phenomena. In particular, in sunspot regions where the strengths of the magnetic field are higher than is normal on the sun's surface, both the chromosphere and the corona have a higher than normal temperature.

The behavior of the chromosphere and the corona is most easily observed by studying the ultraviolet emission from the sun, since in the ultraviolet region the amount of light emitted from the photosphere greatly decreases, whereas the higher temperatures in the chromosphere and corona are responsible for the presence of large numbers of emission lines. The most important emission lines are attributable to hydrogen and helium. In order to understand solar-surface physics in more detail, it is essential to obtain observations of the time variations of these emission lines as indicators of the time variations of behavior in the chromosphere and corona.

The first experiments in this direction were very successfully accomplished by the flight of the first Orbiting Solar Observatory, which was launched on 7 March 1962. It gave data over several months, continuously monitoring a number of different wavelength regions for emission from the sun.

Particularly interesting are the data for the interval 11 through 22 March 1962. At the beginning of this period the sun was in an exceptionally quiet condition, but as the period progressed the sun became more and more active, until on 22 March there was a flare of intensity 3. Experiments revealed that the Lyman alpha line of helium II at 304 angstroms increased in intensity by some 33 percent during the interval, and during the flare itself the intensity of the line increased by an additional 14 percent. The lines of iron XV at 284 angstroms and iron XVI at 335 angstroms also increased in intensity by a factor of 4. At longer wavelengths the Lyman alpha line of hydrogen was observed to increase in intensity by 6.8 percent during the flare.

Very interesting results were also obtained in the x-ray region, 1 to 10 angstroms. During the quiet period a flux was observed which was 360 times the theoretical background radiation which would be obtained from a corona at a temperature of $1.8 \times 10^6$ degrees Kelvin. This indicates that nonthermal processes are present and important in the corona under even the quietest solar conditions.

A continuing series of Orbiting Solar Observatories is planned in which these interesting phenomena can be monitored continuously during future years.

## X-RAYS AND GAMMA RAYS

The space research program is not confined to the discovery of new facts about the solar system. It also gives the astrophysicist an important opportunity to extend his knowledge of more distant parts of space through observations at wavelengths for which photons do not penetrate through the atmosphere. The principal regions involved are the x-ray and gamma ray region, the ultraviolet, the infrared, and long-wavelength radio waves. The early rocket and satellite measurements of x-rays and gamma rays have been particularly interesting to physicists because they suggest several possible new types of phenomena in space.

X-rays and gamma rays can be produced by a variety of high-energy processes. These processes include collisions between high-energy nucleons, which can create neutral pions, which in turn decay to give gamma rays exceeding 50 Mev in energy. Fast electrons can produce X-rays of bremsstrahlung when they pass close to a nucleus. Fast electrons can also collide with photons of visible starlight and increase the energy of the photons up to the x-ray and gamma ray regions. If radioactive nuclei are produced and dispersed in space between the stars, then some of them should emit characteristic gamma ray energies which might be detected. If positrons are produced in dense regions of matter, such as stellar surfaces, then, upon being slowed down and annihilated, they will emit the characteristic gamma rays of 0.51-Mev energy. If neutrons are produced near stellar surfaces and are slowed down

and captured by the overwhelmingly abundant hydrogen that is present, then these will provide characteristic capture gamma rays with an energy of 2.31 Mev. Finally, if objects should exist in space with surface temperatures of some millions of degrees Kelvin, then photons in the x-ray region would be emitted by thermal processes from their surfaces.

Preliminary measurements now exist of the fluxes of x-rays and gamma rays in a number of different energy intervals. A general background of x-rays of energy of a few thousand electron volts was observed in a rocket flight by R. Giacconi, H. Gursky, and F. R. Paolini, of American Science and Engineering, Inc., and by B. B. Rossi of the Massachusetts Institute of Technology. A general background radiation of gamma rays in the region near 1-Mev energy was measured in the Ranger III flight by J. R. Arnold of the University of California (La Jolla), A. E. Metzger, of the California Institute of Technology, and E. C. Anderson and M. A. Van Dilla, of Los Alamos. A small but still significant flux of gamma rays with energies exceeding 50 Mev was observed with the Explorer XI satellite by W. L. Kraushaar and G. W. Clark of the Massachusetts Institute of Technology.

A number of attempts have been made to explain the presence of these background x-rays and gamma rays. Most mechanisms thus far examined appear quantitatively inadequate to explain the observed fluxes. One promising explanation was proposed by J. E. Felten and P. Morrison of Cornell University, who suggested the importance of the inverse Compton effect in which the high-energy electrons present in the cosmic rays collide with photons with energies of the order of 1 electron volt which are emitted from stars. Following such a collision, the energies of the photons can easily be raised to the observed x-ray or gamma ray energies, depending upon the energies of the electrons with which they collide. Calculations by Felten and Morrison were based on this effect (Fig. 9). A flux will be emitted by the outer halo region of our galaxy if the observed flux of high-energy electrons at the earth exists throughout this large outer region of the galaxy. Electrons in the halo fail to account for the observed x-ray and gamma ray fluxes by some 2½ orders of magnitude. However, if the flux of high-energy electrons is the same throughout space as near the earth, then a background radiation some 30,000 times that which would be produced within the galactic halo would be observed. Evidently, such high fluxes of electrons cannot exist throughout space. One percent of such a flux of electrons can be expected to give a background of x-rays and gamma rays which fits the observations very nicely.

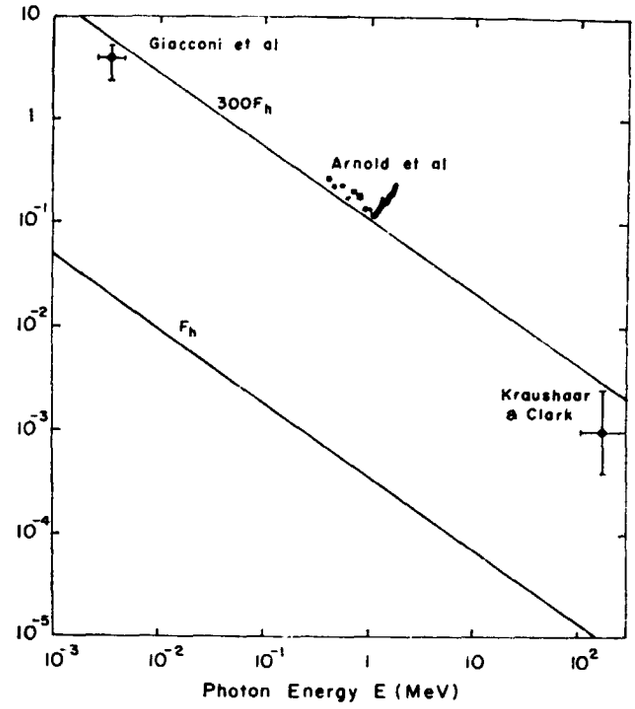


FIGURE 9.—Data on hard photon fluxes in space, compared with theoretical recoil spectra. The data points are as plotted by Felton and Morrison. $F_h$ is the expected contribution from scattering in our galactic halo. The upper curve shows a flux 300 times greater, but still less by two orders of magnitude than the flux that would be obtained if the electrons in our galactic halo extended throughout intergalactic space. [From J. E. Felten and P. Morrison, *Phys. Rev. Letters* 10, 455 (1963)].

Perhaps the most interesting questions concerning the celestial x-rays have been raised through the discovery of discrete sources by Rossi and his colleagues and by H. Friedman. S. Bowyer, T. A. Chubb, and E. T. Byram of the Naval Research Laboratory. Both groups have observed a strong x-ray source in Scorpius which is not coincident with any conspicuous object. Friedman has suggested that this object is a neutron star having a surface temperature of several million degrees, and that the x-rays are due to

thermal emission from the surface layers. Rossi and his colleagues have determined from atmospheric absorption measurements that if the Scorpius source has a thermal spectrum its temperature is approximately $8\times10^6$ degrees Kelvin. Friedman and his colleagues have also observed x-rays from the direction of the Crab Nebula, the remnant of the supernova explosion of A.D. 1054 (Fig. 10).

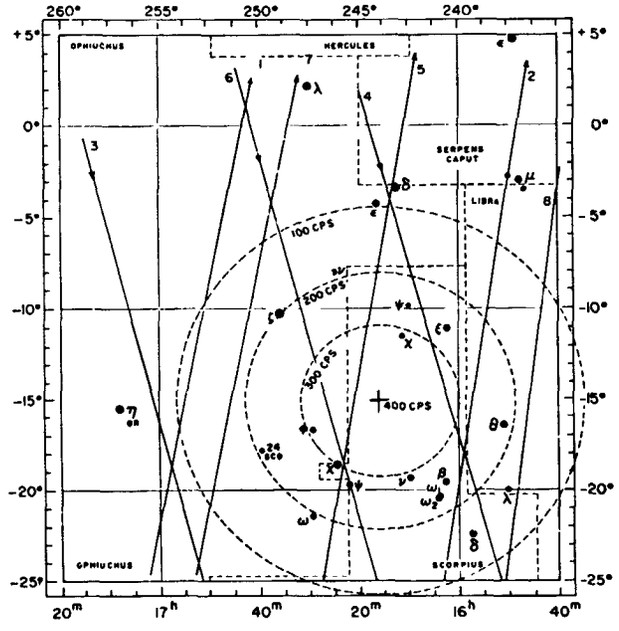


FIGURE 10.—Tracks of eight scans across the Scorpius region. The numbers along the tracks are counts per 0.09-second interval. The dashed circles are best fits to equal-intensity contours and indicate a central intensity peak of 400 count/sec at $\alpha = 16$ hours 15 minutes, $\delta = -15°$. [From S. Bowyer, E. T. Bryam, T. A. Chubb, H Friedman, *Nature* 201, 1307 (1964)].

Neutron stars are hypothetical objects which form one class of degenerate stars, the other class being the degenerate white dwarf stars, which are observed. A typical density for matter in a white dwarf star is $10^6$ g/cm$^3$, and the electrons form a degenerate gas which exerts sufficient pressure to maintain the stars against further contraction. If mass were to be added to such a star, the central region would have to become denser in order to supply the additional pressure required to support the additional mass. There is a relativistic upper limit to the mass of white dwarf stars, but before this limit is reached, the energies of the degenerate electrons have become so high that the nuclei are forced to undergo multiple electron capture reactions, and the nuclei dissolve mainly into neutrons, with only enough protons and electrons left to prevent the neutrons from undergoing their usual mode of decay into electrons and protons.

At $10^{15}$ g/cm$^3$ or more, densities comparable to those in the atomic nucleus, this neutron-rich nuclear matter itself becomes degenerate, and it is expected that stable stars could be constructed of it. Such stars may be formed in the central regions of more massive stars when these stars undergo supernova explosions and blow off most of their mass. Recent work by D. Morton of Princeton, E. E. Salpeter of Cornell, and H. Y. Chiu, S. Tsuruta, and A. G. W. Cameron of the Goddard Institute for Space Studies indicates that the surface temperature of a neutron star is likely to lie below its central temperature by between one and two orders of magnitude. Thus, if such stars are formed with central temperatures over $10^9$ degrees Kelvin, such as would probably be produced in a supernova explosion, then their surface temperatures are likely to be many millions of degrees for several thousand years.

## BIBLIOGRAPHY

### Geodesy

W. M. KAULA, *J. Geophys. Res.* **68**, 5183 (1963).
G. J. F. MACDONALD, *Science* **143**, 921 (1964).

### Meteorology

K. TELEGADS and J. LONDON, *A Physical Model for the Northern Hemisphere for Winter and Summer* (New York Univ. College of Engineering, New York, 1954).
J. LONDON, *A Study of the Atmospheric Heat Balance* (New York Univ. College of Engineering, New York, 1957).
S. FRITZ and H. WEXLER, "Planet Earth as seen from Space," in *Planets and Satellites*, G. P. Kuiper and B. M. Middlehurst Eds. (Univ. of Chicago Press, Chicago, 1961), p. 1.
W. R. BANDEEN, R. A. HANEL, J. LICHT, R. A. Stampfl, W. G. Stroud, *J. Geophys. Res.* **65**, 3169 (1961).
W. NORDBERG, W. R. BANDEEN, B. J. CONRATH V. KUNDE, I. PERSANO, *J. Atmospheric Sci.* **19**, 20 (1962).
D. Q. WARK, G. YAMAMOTO, J. H. LIENESCH, *ibid.*, p. 369.
C. PRABHAKARA and S. I. RASOOL, *International Symposium on Rocket and Satellite Meteorology*, H. Wexler and J. E. Caskey, Jr., Eds. (North-Holland, Amsterdam, 1963), p. 234.
J. S. WINSTON and P. K. RAO, *Monthly Weather Rev.* **91**, 641 (1963).
A. ARKING, *Science* **143**, 569 (1964).
S. I. RASOOL, *ibid.*, p. 567.

### Upper Atmosphere

M. NICOLET, *J. Geophys. Res.* **66**, 2263 (1961).
R. E. BOURDEAU, E. C. WHIPPLE, JR., J. L. DONLEY, S. J. BAUER, *ibid*, **67**, 467 (1962).
I. HARRIS and W. PRIESTER, *J. Atmospheric Sci.* **19**, 286 (1962).
———, *J. Geophys. Res.* **67**, 4585 (1962).
———, *ibid.* **68**, 5891 (1963).
L. G. JACCHIA, in *Space Research* (vol. 3 of *International Space Science Symposium*), W. Priester Ed. (North-Holland, Amsterdam, 1963), p. 3.
L. G. JACCHIA and J. SLOWEY, *J. Geophys. Res.* **69**, 905 (1964).

### Magnetosphere

S. F. SINGER, *Trans. Am. Geophys. Union* **38**, 175 (1957).
J. A. VAN ALLEN, G. H. LUDWIG, E. C. RAY, C. E. MCILWAIN, *Jet Propulsion* **28**, 588 (1958).
B. J. O'BRIEN, *J. Geophys. Res.* **67**, 1227 (1962).
———, *ibid.*, p. 3687.
———, *Space Sci. Rev.* **1**, 415 (1962).
S. AKASOFU and S. CHAPMAN, *J. Geophys. Res.* **66**, 1321 (1963).

### Magnetopause

L. A. FRANK, J. A. VAN ALLEN, E. MACAGNO, *ibid.* **68**, 3543 (1963).
L. J. CAHILL and P. G. AMAZEEN, *ibid.*, p. 1835.
A. BONETTI, H. S. BRIDGE, A. J. LAZARUS, B. ROSSI, F. SCHERB, *ibid.*, p. 4017.
H. R. ANDERSON, *Science* **139**, 42 (1963).
E. PARKER, *Interplanetary Dynamical Processes* (Interscience, New York, 1963).

### Atmosphere of Venus

C. H. MAYER, T. P. MCCULLOUGH, R. M. SLOANAKER, *Astrophys. J.* **127**, 1 (1958).
D. E. JONES, *Planetary Space Sci.* **5**, 166 (1961).
S. I. RASOOL, *Am. Inst. Aeron. Astronaut, J.* **1**, 6 (1963).
S. C. CHASE, L. D. KAPLAN, G. NEUGEBAUER, *J. Geophys. Res.* **68**, 6157 (1963).
R. T. BARATH, A. H. BARRETT, J. COPELAND, D. E. JONES, E. A. LILLY, *Astron. J.* **69**, 49 (1964).

### Exploration of the Moon

N. P. BARABASHOR, A. A. MIKHAILOV, Y. N. LIPSKIY, *Atlas of the Other Side of the Moon* (Pergamon, London, 1961).
S. S. DOLGINOV, E. G. EROSHENKO, L. I. ZHUZGOV, N. V. PUSHKOV, "A study of the magnetic field of the Moon," in *The Moon*, Z. Kopal and Z. V. Mikhailov, Eds. (Academic Press, New York, 1962).

### Solar Physics

D. E. OSTERBROCK, *Astrophys. J.* **134**, 347 (1961).
W. A. WHITTAKER, *ibid.* **137**, 914 (1963).
J. C. LINDSAY, *Trans. Am. Geophys. Union* **44**, 722 (1963).
D. W. MOORE and E. A. SPIEGAL, *Astrophys. J.* **138**, 48 (1964).

### X-rays and Gamma Rays

W. L. KRAUSHAAR and G. W. CLARK, *Phys. Rev. Letters* **8**, 106 (1962).
R. GIACCONI, H. GURSKY, F. R. PAOLINI, B. R. ROSSI, *ibid.* **9**, 439 (1962).
J. ARNOLD, A. E. METZGER, E. C. ANDERSON, M. A. VAN DILLA, *J. Geophys. Res.* **67**, 4878 (1962).
J. E. FELTEN and P. MORRISON, *Phys. Rev. Letters* **10**, 453 (1963).
S. BOWYER, E. T. BYRAM, T. A. CHUBB, H. FRIEDMAN, *Nature* **201**, 1307 (1964).
D. MORTON, *ibid.*, p. 1308.
H. Y. CHIU and E. E. SALPETER, *Phys. Rev. Letters* **12**, 413 (1964).
———, *Ann. Phys. N.Y.* **26**, 364 (1964).

# $g_J$ OF THE $(5s5p)\ ^3P_1$ LEVEL OF Cd AND THE $(6s6p)\ ^3P_1$ LEVEL OF Hg BY HIGH-FIELD DOUBLE RESONANCE*

R. KOHLER† AND P. THADDEUS

An optical double resonance experiment at high magnetic fields has determined $g_J$ for the $(5s5p)\ ^3P_1$ level of the even cadmium isotopes to be 1.499846(13), and $g_J$ for the $(6s6p)\ ^3P_1$ level of the even mercury isotopes to be 1.486094(8). Resonance was observed in all cases at a frequency of exactly 24 Gc/sec, and magnetic fields near 11 430 G. At these high fields the Zeeman energy is of the order of $10^{-3}$ of the fine structure separations of the triplet terms of the $(nsnp)$ configuration, and the $\Delta m = \pm 1$ transitions are split by 9.51(7) G for Cd, and by 2.99(7) G for Hg. This splitting represents several linewidths in the case of Hg, and many linewidths in that of Cd. The average field of the two $\Delta m = \pm 1$ transitions, however, determines $g_J$ to high precision independent of second-order corrections.

## I. INTRODUCTION

The double resonance[1,2] and level crossing techniques[3,4] have been used in extensive studies of hyperfine structure (hfs) in the lowest lying $(nsnp)\ ^3P_1$ level of mercury,[5–11] cadmium,[12–17] and zinc.[18–20] Sufficient sensitivity has been obtained to allow the determination of the hfs, and consequently the nuclear moments, of radioactive isotopes with half-lives as short as a few hours.[15,16]

While the level crossing method has the great advantage in practice of doing away with all radio-frequency and microwave apparatus, aside from nuclear magnetic resonance (NMR) magnetometers, it allows a determination of only the ratio of the hyperfine interaction constants to $g_J$, and an independent measurement of $g_J$ must therefore be made to determine these constants to high precision.[16] While $g_J$ can of course be measured by double resonance in the odd isotopes at various fields, or by observing both double resonance and level crossings,[9,13] it is most directly found from double resonance in the even isotopes of the element.

In principle, the higher the frequency and the greater the magnetic field, the more accurate the determination of $g_J$. In a series of experiments designed to measure $g_J$ of the lowest lying $^3P_1$ level of zinc, cadmium, and mercury to high precision (it is $\frac{3}{2}$ on the basis of $LS$ coupling), we therefore chose to work at a frequency of 24 Gc/sec, and a field near 11 400 G—approaching the limit attainable with the available 12-in. electromagnet when operating with a gap of 2 in.

The $\Delta m = \pm 1$ transitions for these states coincide at low magnetic fields, but at high fields interaction with the neighboring fine structure terms of the configuration (and to a negligible degree with levels of other configurations) splits the resonances into two well-resolved components, as shown in Fig. 1. Precise measurement of the splitting of the two resonances therefore serves also as a check on the theory of the second-order fine structure interactions.

## II. THEORY

### Signal Strengths

It is well known that double resonance can be observed between hyperfine or Zeeman states excited by resonance radiation in a variety of ways. Population differences can be produced by resonance radiation that is polarized, nonisotropic, or confined to a narrow wavelength interval,

*Published in *The Physical Review*, 134 (5A): A 1204-A1209, June 1, 1964 work supported by the Joint Services (the U.S. Army, the U.S. Navy, Officers of Naval Research, and the U.S. Air Force Office of Scientific Research), and by the U.S. Office of Naval Research under, contract Nonr-266645).

†Columbia University, New York, and New York University, New York.

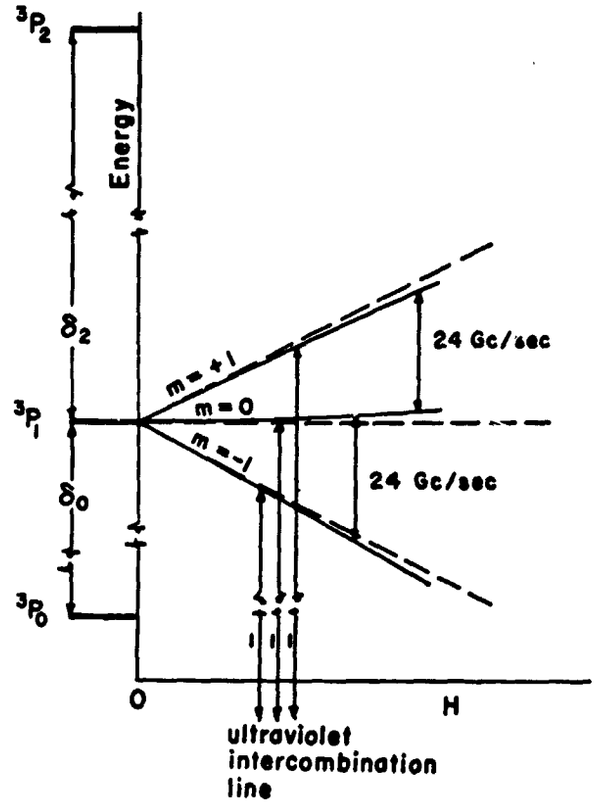


FIGURE 1.—Zeeman effect of the $^3P_1$ state of the even isotopes of zinc, cadmium and mercury, showing the effect of second-order interaction with the other fine structure levels of the configuration.

and all of these methods have been put to practical use.[1,2,6]

To dispense with ultraviolet polarizers, we relied on the latter two effects and used the exeprimental geometry shown in Fig. 2. Only lamps containing the natural metal were used. In the case of cadmium, the Zeeman shift of the $m=\pm 1$ states at 11 400 G is somewhat greater than the isotope shifts and the hyperfine splittings of the natural cadmium isotopes (and the optical Doppler width),[14] and only the $m=0$ state is therefore illuminated. It is then readily calculated in terms of the transition probabilities that at double resonance the resonance fluorescence at 90° is reduced by a factor of $\frac{1}{4}$.

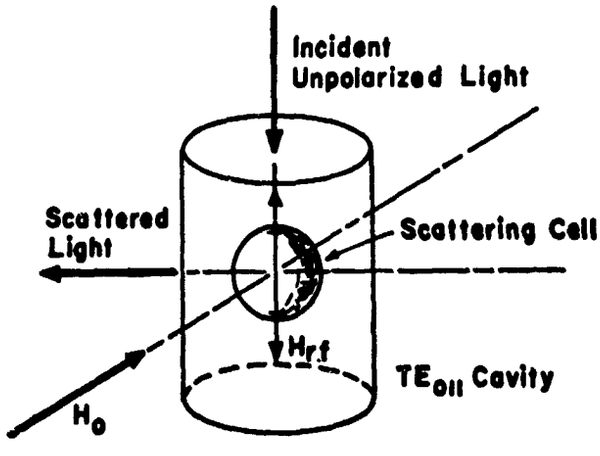


FIGURE 2.—Geometry of the experiment.

For mercury, on the other hand, the isotope shifts and hyperfine splittings are comparable to the Zeeman shift of the $\Delta m=\pm 1$ states,[21] and we must rely mainly on the nonisotropic nature of the resonance radiation. In the least favorable case, on the assumption that all three states are equally illuminated, it is calculated that the resonance fluorescence at 90° is reduced by a factor of 1/12 at double resonance.

### Linewidths

Several effects determine the linewidth and shape of the double resonance effect, and have all been considered in detail elsewhere. Most fundamentally, the natural lifetime $\tau$ of the state gives a Lorentzian line shape with full width at half-maximum of $g_J\mu_0\Delta H_n=\Delta\nu_n=1/\pi\tau$.[1,6] The natural linewidths for zinc, cadmium, and mercury calculated from the lifetimes in Table I are listed in Table II.

If the rf magnetic field is strong enough to induce many transitions during the lifetime $\tau$, the resonance will be power broadened.[1,6] In practice the rf power is always attenuated to the point where this effect is small, and it will therefore not be considered further.

Double resonance owes its high resolution to the fact that Doppler broadening is determined by the radio rather than the optical frequency. At microwave frequencies, however, the Doppler linewidth $\Delta\nu_d=(\nu/c)\times(8NkT\ \ln 2/M)^{1/2}$, where $N$ is Avogadro's number and $M$ is the molecular weight, may not be altogether negligible, as is shown in Table II.

Inhomogeneity of the magnetic field over the scattering sample contributes a linewidth $\Delta H_m$ that can often be reduced by careful alignment of the pole pieces. We found, however, that as our magnetic field approached 11 400 G, its homogeneity fell off rapidly due presumably to uneven

TABLE I.—*Atomic constants of Zn, Cd, and Hg.* $\lambda(^3P_1)$ *and* $\lambda(^1P_1)$, *the wavelengths of the resonance lines of the lowest lying* ($n$snp) *configuration, and* $\delta_2$ *and* $\delta_0$, *the fine structure separations of the triplet terms, are taken from C. E. Moore.*[a]

| Atom | $\tau$ ($^3P_1$) (sec) | $\tau$ ($^1P_1$) (sec) | $\lambda$ ($^3P_1$) (Å) | $\lambda$ ($^1P_1$) (Å) | $\delta_2$ ($cm^{-1}$) | $\delta_0$ ($cm^{-1}$) |
|---|---|---|---|---|---|---|
| Zn | $2.00(20)\times10^{-8}$ [b] | $1.38(5)\times10^{-9}$ [c] | 3077 | 2139 | 388.927 | 190.082 |
| Cd | $2.39(4)\times10^{-6}$ [d] | $1.66(5)\times10^{-9}$ [e] | 3261 | 2288 | 1170.866 | 542.113 |
| Hg | $1.18\times10^{-7}$ [f] | $0.3$–$1.6\times10^{-9}$ [g] | 2537 | 1849 | 4630.677 | 1767.220 |

a. C. E. Moore, *Atomic Energy Levels* (U.S. Government Printing Office, 1949–58).
b. See Ref. 19.
c. See Ref. 20.
d. See Ref. 23.
e. A. Lurio and R. Novick, Phys. Rev. 134, A608 (1964).
f. See the last paper of Ref. 22.
g. A. C. G. Mitchell and M. W. Zemansky, *Resonance Radiation and Excited Atoms* (Cambridge University Press, Cambridge, 1961), 2nd ed., p. 147

TABLE II.—*Theoretical and observed linewidths in milligauss.* $\Delta H_{calc}$ *is taken as the sum of the natural, Doppler, magnetic field, and wall collision widths.*

| Atom | $\Delta H_n$ | $\Delta H_d$ | $\Delta H_m$ | $\Delta H_w$ | $\Delta H_{calc}$ | $\Delta H_{obs}$ |
|---|---|---|---|---|---|---|
| Zn | 8 | 29 | 50 | 37 | 124 | ------- |
| Cd | 68 | 18 | 50 | 26 | 162 | 175 |
| Hg | 1295 | 10 | 50 | 15 | 1370 | 1300 |

saturation. We estimate that $\Delta H_m = 50$ mG over a 5-mm sample.

Collisions with the walls of the scattering cell effectively shorten the lifetime of the excited state, and contribute a wall-quenching linewidth $\Delta\nu_w$ which depends mainly on the size of the cell. We have made estimates of $\Delta\nu_w$ in Table II on the assumption that $\Delta\nu_w = 1/\pi\tau_w = (3NkT/M)^{1/2}/\pi a$, where $\tau_w$ is the time for an atom at the temperature of the optimum vapor pressure to travel the radius $a$ of the scattering cell.

Collisions between the scattering atoms will broaden the resonance at high vapor densities, while multiple scattering of the optical photons may actually decrease the linewidth (coherence narrowing).[22,23] Both of these effects may be avoided by working at low enough temperatures.

Quenching by atomic or molecular contaminants in the scattering cell, on the other hand, may contribute a linewidth $\Delta\nu_c$ which cannot be easily estimated in advance, and cannot be reduced once the cell is made. We shall see that possibly except for zinc we found no evidence of quenching of this kind.

### Zeeman Effect

Interaction with the two neighboring triplet levels of the ($nsnp$) configuration contributes a term which is of the order of the (Zeeman energy)$^2$/(fine structure energy) $\approx 10^{-3}\times$(Zeeman energy) to the energy of the $m=\pm1$, 0 states. More precisely, from perturbation theory,[24]

$$W_{+1} = g_J\mu_0 H - \frac{\alpha^2(g_S-g_L)^2\mu_0^2H^2}{4\delta_2}, \qquad (1a)$$

$$W_0 = \alpha^2(g_S-g_L)^2\mu_0^2H^2\left(\frac{2}{3\delta_0}-\frac{1}{3\delta_2}\right), \qquad (1b)$$

$$W_{-1} = -g_J\mu_0 H - \frac{\alpha^2(g_S-g_L)^2\mu_0^2H^2}{4\delta_2}, \qquad (1c)$$

where $\mu_0$ is the Bohr magneton, $\alpha$ is the mixing coefficient which indicates the departure from Russell-Saunders coupling, and $\delta_0$ and $\delta_2$ are the magnitude of the fine structure separations. We adopt the convention that $g_J$ is positive. Interaction with the singlet term of the configuration, and with terms of other configurations, contributes less than one part in $10^6$ to the energy, and third-order terms vanish.

For double resonance at a fixed frequency $\nu_{dr}$, the field $H^+$, for which $m=0\rightarrow+1$, is greater than $H^-$, for which $m=0\rightarrow-1$. To within one part in $10^6$ we find from Eq. (1) that in terms of the average field $\bar{H}=(H^++H^-)/2$

$$g_J = h\nu_{dr}/\mu_0\bar{H}. \qquad (2a)$$

This becomes

$$g_J = g_p(\nu_{dr}/\nu_p), \qquad (2b)$$

if the field is measured in terms of the NMR frequency $\nu_p$ of protons. $g_p$ is defined such that the proton magnetic dipole-moment operator is $\mathbf{u}_p = g_p \mu_0 \mathbf{I}_p$. For protons in mineral oil we will take $g_p = 3.041978(3) \times 10^{-3}$, from the average of the measurements of $g_s/g_p$ for protons in mineral oil of Koenig, Prodell, and Kusch,[25] and Beringer and Heald,[26] and the recent measurement of $g_s$ of Wilkinson and Crane.[27]

For the separation of the two resonances we find that

$$\Delta H = H^+ - H^- = \frac{\alpha^2 (g_S - g_L)^2 \mu_0 H^2}{g_J} \left( \frac{4}{3\delta_0} - \frac{1}{6\delta_2} \right), \quad (3a)$$

where $\delta_0$ and $\delta_2$ are in ergs. When the field is measured in terms of the proton frequency,

$$\Delta \nu_p = \frac{\alpha^2 (g_S - g_L)^2 \bar{\nu}_p^2}{c g_p g_J} \left( \frac{4}{3\delta_0} - \frac{1}{6\delta_2} \right), \quad (3b)$$

where $\delta_0$ and $\delta_2$ are in $\text{cm}^{-1}$.

## III. EXPERIMENT

A schematic illustration of the experiment is shown in Fig. 3. Resonance radiation from an rf lamp was focused with a quartz condensing lens onto a small cell containing cadmium or mercury vapor. The cell was located within a microwave cavity between the poles of a 12-in. electromagnet. This cavity in the case of cadmium also served as an oven to control the vapor density in the scattering cell.

Resonance fluorescence of the atomic vapor was observed with a photomultiplier at an angle of 90°. A light pipe allowed the photomultiplier to be removed from the magnetic field without loss of solid angle. A glass color filter effectively removed all scattered light except the resonance radiation.

The microwave cavity was tuned to resonance with a klystron kept at a frequency of exactly 24 Gc/sec, the field being modulated at a low audio frequency, typically 37 cps, by small rectangular coils attached to the pole faces. At resonance, the light scattered at 90° changed in intensity, and the signal from the photomultiplier was displayed on an oscilloscope.

Field measurements were made with an NMR magnetometer, whose small mineral-oil probe was located as close to the scattering cell as possible. The offset correction was measured with a second probe and oscillator.

### Optics

The cadmium lamp used consisted of a small quartz cylinder, 1.5 cm diameter and 2.5 cm long, carefully baked out, purged with an rf discharge in argon and then filled with a small amount of distilled natural cadmium and about 1 torr of spectroscopically pure argon. It was identical to the lamps previously used in this laboratory for several level crossing and double resonance experiments.[13,14] It was excited by being placed, without cooling, in the tank coil of a Hartley oscillator operating in the vicinity of 30 Mc/sec. Experience has shown that these lamps furnish about 1 W (total output over $4\pi$ solid angle) in the 3261-Å intercombination line.

For mercury an Osram lamp was used that had the outer jacket removed, and was cooled with a gentle air stream to eliminate self-reversal. It too was excited with a radiofrequency oscillator operating near 30 Mc/sec and was run at the same

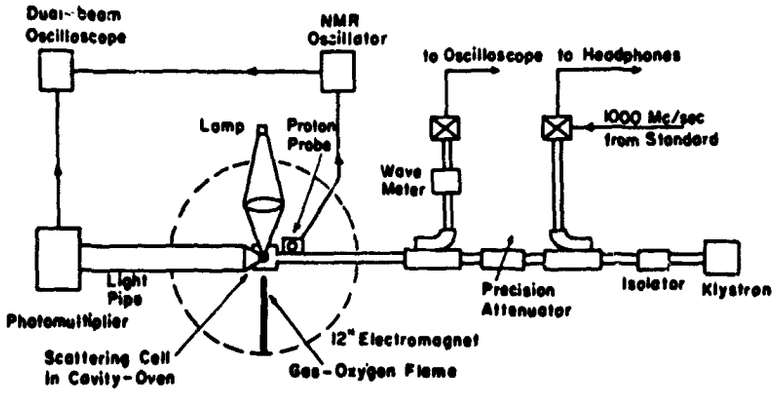


FIGURE 3.—Schematic diagram of the experiment.

current specified by the manufacturer for operation at 60 cps.

The scattering cells were designed to have as small an effect as possible on the electrical properties of the microwave cavity. They were very thin walled quartz bubbles, 5 mm in diameter, blown at the end of a length of 2 mm outer diameter quartz tubing, which, when cut off at a length of 5 cm, served as the tail of the cell. These cells were baked out at 1000°C for a day or longer, purged with a Tessla coil for several minutes, and filled with a small quantity of distilled metal, all under very high vacuum (less than $10^{-8}$ Torr for the last stage). The light pipe used to remove the RCA-1P28 photomultiplier from the vicinity of the magnet was a glass tube aluminized on the inside, 1 in. in diameter, and about 2½ ft. long. To pass the 3261-Å line of cadmium, or the 2537-Å line of mercury, and exclude other scattered light, a Schott UG-11 or UG-5 colored glass filter was cemented to the end of the light pipe.

### Microwave Cavity and Oven

The $TE_{011}$ cylindrical aluminum cavity oven used for cadmium is shown in Fig. 4. It had an internal diameter of 1.4 cm, a $Q$ of about 4000 when loaded with the cell, and was coupled about 50 percent to the microwave line. A hole 6 mm in diameter at the top of the cavity, below cutoff at 24 Gc/sec, allowed illumination of the scattering cell. Small slits in the side of the cavity, also below cutoff, allowed the light scattered at 90° to be observed.

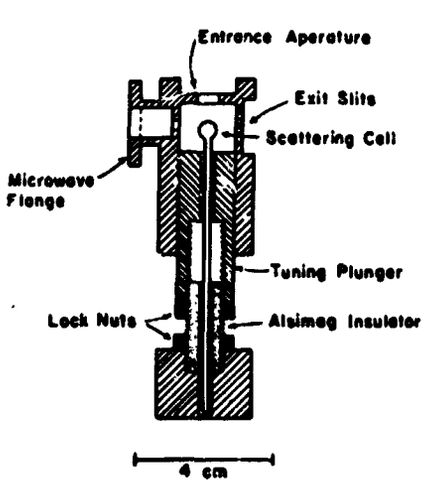


FIGURE 4.—The cavity oven used for cadmium.

The waveguide connecting the cavity to the microwave line was thinned to provide thermal insulation. The cavity was heated with a small gas-oxygen flame, the temperature being monitored by a small copper-constantin thermocouple. Thermal regulation was provided by a Honeywell R7086A thermocouple-activated regulator that played a stream of cool air against the cavity when the thermocouple voltage exceeded a preset figure.

The long tail of the scattering cell passed through an aluminum tuning plunger, and an insulating refractory section, and terminated in the aluminum cylinder shown in the bottom of Fig. 4. This cylinder was heated with another gas-oxygen flame, and its temperature regulated with a separate regulator unit. It was always kept at least 40°C cooler than the oven, and consequently determined the cadmium vapor pressure.

For mercury, in an attempt to gain a higher $Q$, a thin-walled copper cavity was used that was kept at room temperature. With a brass tuning plunger, a loaded cavity $Q$ of about 8000 was obtained. The tail of the cell was usually kept at room temperature, but could also be cooled in an ice bath.

### Microwave Measurements

The Varian VA-98E klystron used in this experiment is an extremely stable tube that oscillates only over the region from 23.6 to 24.4 Gc/sec, with an average output of about 30 mW. To achieve maximum stability, we used batteries for both the heater and repeller voltages, and a well-regulated power supply for the beam voltage. No electronic stabilization of the tube was employed. During a run it was kept manually to within an audiobeat not of the 24th harmonic of the 1000 Mc/sec output of the laboratory's frequency standard. This represents a stability of about one part in $10^6$, and is therefore quite adequate for the purpose of this experiment.

### Field Measurements

The proton probe of the NMR magnetometer used for field measurements consisted of a small spherical cavity filled with mineral oil drilled in the center of a teflon spool. The inductance for a transistorized marginal oscillator was wound on the spool which was then placed in a brass or copper shield. It was located as close as possible

to the scattering cell (about 4 cm away), and the output signal of the marginal oscillator was displayed with the double resonance signal on a dual beam oscilloscope. The NMR frequency, near 48 Mc/sec, was measured with a Hewlett-Packard direct reading frequency counter whose reference frequency was taken from the laboratory standard. The offset correction was measured by removing the oven, placing a second probe driven by another identical oscillator at the scattering cell location, and measuring the beat frequency ($\approx 4$ kc/sec) of the two oscillators with the frequency counter. A pentode mixer gave the difference frequency without excessive cross coupling of the two oscillators.

The offset correction was found to depend on the field history, and a correction was therefore made after each field measurement.

## IV. RESULTS

For cadmium the onset of power broadening was observed when about 5 mW of microwave power were incident on the cavity, in rough accord with our preliminary estimates based on the lifetime of the $(5s5p)\ ^3P_1$ level, and the electrical properties of the cavity. With oscilloscope display giving a detector bandwidth of about 1 kc/sec we succeeded in obtaining a signal-to-noise ratio of about 30 for both the $\Delta m = \pm 1$ double resonance transitions [Fig. 5(a)]. The observed linewidth (Table II) shows no evidence of contamination broadening.

The shorter lifetime of the $(6s6p)\ ^3P_1$ level of mercury was found to require for maximum signal strength somewhat higher rf fields than the VA-98E klystron and aluminum cavity could provide. A higher $Q$ copper cavity subsequently used for mercury just allowed the level of power broadening to be reached. The observed resonances [Fig. 5(b)] had a signal-to-noise ratio of about 5. It should be recalled that because of the illumination of the $m = \pm 1$ states the mercury signals are expected to be about 3 times weaker than those of cadmium. The additional discrepancy in signal strength can be attributed to the somewhat weaker output of the mercury lamp. The linewidth was in agreement with expectations (Table II).

We had originally hoped to extend our measurements to the $(4s4p)\ ^3P_1$ level of the even zinc isotopes, although we were aware that the long lifetime of this level would produce serious quenching of the excited state atoms against the wall of the scattering cell. We, in fact, made several runs using narrow-band phase-sensitive detection with time constants as long as 5 sec, but failed to detect either double resonance or fluorescence. Beyond the quenching and broadening mechanisms listed in Table II, we are inclined to attribute this to contaminants produced by the high cell temperatures required for zinc.

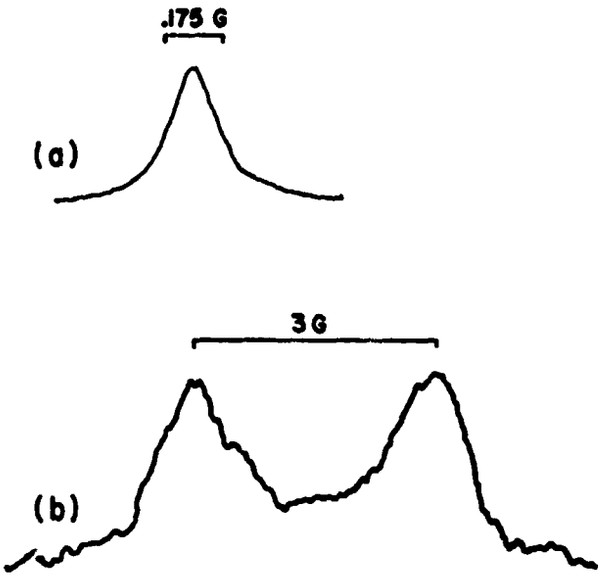


FIGURE 5.—Oscilloscope traces of the double resonance signals: (a) is one of the $\Delta m = \pm 1$ cadmium transitions, (b) is the sum of these two transitions for mercury, split by the second-order interactions.

The magnetic fields at which double resonance was observed for cadmium and mercury are listed in Table III. From Eq. (2) we then find for the $(5s5p)\ ^3P_1$ level of cadmium that

$$g_J = 1.499846(13) = \tfrac{3}{2} - 154(13) \times 10^{-6},$$

in good agreement with the values $g_J^{111} = \frac{3}{2} - 171(9) \times 10^{-6}$ and $g_J^{113} = \frac{3}{2} - 157(9) \times 10^{-6}$ found for the odd isotopes from level crossings[13] and zero-field double resonance[12] (we have here recalculated $g_J^{111}$ and $g_J^{113}$ taking $g_p = 3.041978 \times 10^{-3}$).

Likewise, we find for the $(6s6p)\ ^3P_1$ level of mercury that

$$g_J = 1.486094(8) = \tfrac{3}{2} - 13\ 906(8) \times 10^{-6}.$$

This is in reasonably good agreement with a recent preliminary value of $g_J^{199} = 1.486118(16)$ found by

Smith,[28] who considered both his own level-crossing study of $Hg^{199}$, and the results of low-field double resonance.[11] It is also in agreement with the less precise value 1.486350(300) found by Dodd[29] from double resonance in the even isotopes. It is in disagreement, however, with the self-consistent values $g_J^{199}=1.486147(10)$ and $g_J^{201}=1.486156(18)$ calculated by Kaul in his thesis[9] from his study of level crossings, and the low-field double resonance results.[6,11]

For cadmium the known lifetimes of the $^3P_1$ and $^1P_1$ levels allow $\alpha$ to be calculated directly from the relation

$$\frac{\beta^2}{\alpha^2}=\frac{\tau(^1P_1)\lambda^3(^3P_1)}{\tau(^3P_1)\lambda^3(^1P_1)}, \tag{4}$$

and the normalization condition $\alpha^2+\beta^2=1$. Using the lifetimes and wavelengths of Table I we find that $\alpha=0.99900(5)$ and $\beta=-0.0448(11)$. The phase convention of Condon and Shortley is used to determine the sign of $\beta$.[24] From Eq. (3) the splitting of the resonances is then calculated to be $\Delta\nu_p=40.245(4)$ kc/sec, or $\Delta H=9.452(1)$ G. This is seen from Table III to be 0.255 kc/sec lower than the experimental value, but just within the experimental uncertainty.

Knowing $\alpha$ and $\beta$ independently from the lifetimes, we are in a position to determine the sum of the relativistic and diamagnetic corrections to $g_J$. Theoretically

$$g_J=\tfrac{1}{2}\alpha^2(g_L+g_S)+\beta^2 g_L+\Delta g_{\rm rel}+\Delta g_{\rm diam}. \tag{5}$$

Taking $g_L=1$,[30] and $g_S=2.002319$,[27] we find that[31]

$$\Delta g_{\rm rel}+\Delta g_{\rm diam}=-318(40)\times 10^{-6}.$$

Since for mercury the lifetime of the $^1P_1$ level is poorly known, the sum of the relativistic and diamagnetic corrections cannot be found. Instead, however, we may use Eq. (5) neglecting these corrections to obtain $\alpha=0.9849(4)$ and $\beta=-0.1733(24)$. We then calculate from Eq. (3) the splitting of the mercury resonances be to $\Delta\nu_p=12.468(8)$ kc/sec or $\Delta H=2.928(2)$ G. It is interesting to note that this is lower than the experimental value by 0.252 kc/sec, nearly the same as in the case of cadmium. Since the measurement of the splitting $\Delta H$ is not susceptible to the same systematic errors as the absolute field measurements, the uncertainties in $\Delta H$ given in Table III must be considered quite conservative, and this discrepancy may be significant.

## V. DISCUSSION

While an accuracy of about one part in $10^5$ has been attained in the field measurements reported in this experiment, for cadmium this represents an uncertainty of one linewidth, and can certainly be improved upon. In retrospect we believe that higher precision could have been realized at somewhat lower frequencies and fields. Most of our uncertainty can be ascribed to the inhomogeneities which crept in as the field approached 11 400 G. Had we worked at half the present field and frequency, for example, the field would have been

TABLE III.—*Measured fields for double resonance in cadmium and mercury, in terms of the NMR frequency of protons in mineral oil, and in G, taking for $1/2\pi\times$ the gyromagnetic ratio 4.25760(4) kc/sec-G.*[a] *A series of runs were made for each atom, care being taken to randomize systematic errors from run to run by slightly relocating the cell and probe within the field, reversing the polarity of the magnet, varying the modulation frequency, changing the vapor pressure, reassembling the optics, etc. The uncertainties quoted are 3 times the standard deviation of the means of the individual runs.*

| Atom | $H^+$ | $H^-$ | $\Delta H$ |
|---|---|---|---|
| Cd | 48 696.91(66) kc/sec | 48 656.40(45) kc/sec | 40.50(28) kc/sec |
| | 11 437.64(19) G | 11 428.13(15) G | 9.51(7) G |
| Hg | 49 133.45(35) kc/sec | 49 120.73(38) kc/sec | 12.72(26) kc/sec |
| | 11 540.17(14) G | 11 537.19(14) G | 2.99(7) G |

[a] R. L. Driscoll and P. L. Bender, Phys. Rev. Letters 1, 413 (1958).

considerably more uniform, the scattering cell would have been twice as large (and easier to construct and more durable), the Doppler and wall collision linewidths halved, and the exit and entrance apertures increased in area by four. It is probable that under these circumstances we would have had sufficient signal strength to observe double resonance in zinc, and that the cadmium resonance could have been measured to a small fraction of its width.

## ACKNOWLEDGMENTS

We wish to express our thanks to the staff of the Columbia Radiation Laboratory for their invaluable assistance, and to Professor R. Novick for his advice and support. We would like especially to acknowledge the assistance of H. Feldman, who collaborated with us during a part of the cadmium experiment. Finally, Raymond Eisenstark assisted us liberally during all phases of this work as data taker and calculator, and T. Psaropulos kindly prepared the drawings.

## REFERENCES

1. J. Brossel and F. Bitter, Phys. Rev. 86, 308 (1952).
2. G. W. Series, Rept. Progr. Phys. 22, 280 (1959).
3. F. D. Colegrove, P. A. Franken, R. R. Lewis, and R. H. Sands, Phys. Rev. Letters 3, 420 (1959).
4. P. A. Franken, Phys. Rev. 121, 508 (1961).
5. J. N. Dodd, Proc. Phys. Soc. (London) 77, 669 (1961).
6. R. H. Kohler, Phys. Rev. 121, 1104 (1961).
7. H. R. Hirsch, J. Opt. Soc. Am. 51, 1192 (1961).
8. C. Brot, J. Phys. Radium 22, 412 (1961).
9. R. D. Kaul, Ph.D. thesis, Case Institute of Technology, 1963 (unpublished).
10. W. Smith, Ph.D. thesis, Massachusetts Institute of Technology, 1963 (unpublished).
11. C. V. Stager, Phys. Rev. 132, 275 (1963).
12. R. F. Lacey, Ph.D. thesis, Massachusetts Institute of Technology, 1959 (unpublished).
13. P. Thaddeus and R. Novick, Phys. Rev. 126, 1774 (1962).
14. M. N. McDermott and R. Novick, Phys. Rev. 131, 707 (1963).
15. F. W. Byron, M. N. McDermott, and R. Novick, Phys. Rev. 132, 1181 (1963).
16. P. Thaddeus and M. N. McDermott, Phys. Rev. 132, 1186 (1963).
17. M. N. McDermott, R. Novick, B. W. Perry, and E. Saloman, Phys. Rev. 134, B25 (1964).
18. V. K. Böckman, H. Krüger, and E. Rechnagel, Ann. Physik 20, 250 (1957).
19. F. W. Byron, M. N. McDermott, R. Novick, B. W. Perry, and E. Saloman, Phys. Rev. 134, A47 (1964).
20. A. Landman and R. Novick, Phys. Rev. 134, A56 (1964).
21. F. Bitter, S. P. Davis, B. Richter, and J. E. R. Young, Phys. Rev. 96, 1531 (1954).
22. J. P. Barrat, J. Phys. Radium 20, 541, 633, 657 (1959).
23. F. W. Byron, M. N. McDermott, and R. Novick, Phys. Rev. 134, A615 (1964).
24. A. Lurio, M. Mandel, and R. Novick, Phys. Rev. 126, 1758 (1962).
25. S. H. Koenig, A. G. Prodell, and P. Kusch, Phys. Rev. 88, 191 (1952).
26. R. Beringer and M. A. Heald, Phys. Rev. 95, 1474 (1954).
27. D. T. Wilkinson and H. R. Crane, Phys. Rev. 130, 852 (1963).
28. W. Smith (private communication).
29. J. N. Dodd, Proc. Phys. Soc. (London) 78, 65 (1961).
30. This neglects effects due to nuclear motion, which are expected to contribute a correction to $g_L$ of the order of a few ppm. See A. Abragam and J. H. Van Vleck, Phys. Rev. 92, 1448 (1953).
31. The theoretical values of $g_J=\frac{3}{2}-18(80)\times10^{-6}$ and $\Delta g_{\text{rel}}+\Delta g_{\text{diam}}=-135(81)\times10^{-6}$ given in Ref. 13 are in error. These values were calculated from Eq. (13), Ref. 13 on the basis of outdated lifetimes—not those given in Table I, Ref. 13. To the best of our knowledge, the remaining results in that work are self-consistent.

N64-24084

# ORIGIN OF DIAMONDS IN THE URELITES*

MICHAEL E. LIPSCHUTZ†

Diamonds in two of the three diamantiferous stony meteorites (ureilites) show a pronounced crystallographic orientation. This suggests that the diamonds were produced by shock during breakup of the meteorite's parent body rather than by high gravitational pressure.

In previous papers (*1*, *2*) on the origin of meteoritic diamonds (Table 1) I concluded that the diamonds in the Canyon Diablo iron were formed by conversion of graphite as a result of the shock of impact of the meteoroid with the earth. The subsequent finding of coesite and stishovite (*3*) in the sandstone around the Canyon Diablo crater indicates shock pressures of at least 100 kbar. The successful shock-induced conversion of graphite to diamond in the laboratory (*4*) also lends strong support to the shock origin of Canyon Diablo diamonds.

Diamonds also occur in a rare type of stony meteorite, the ureilites (*5*). These meteorites, however, are too small (Table 1) to have hit the earth at speeds much greater than terminal velocity, and this implies that the diamonds could not have been formed during the impact of the ureilites with the earth. I therefore suggested (*1*) that the diamonds in these meteorites were formed by some preterrestrial shock—probably the one occurring during catastrophic breakup of the ureilites' parent body. In this paper I report on an x-ray study of meteoritic and synthetic diamonds which was undertaken to examine this suggestion.

TABLE 1.—*Identification of diamonds in ureilites and the phases identified in individual diamonds.*

| Meteorite | Recovered mass (kg) | Phases identified |
|---|---|---|
| Dyalpur (*21*) | 0·26 | Diamond, graphite, α-iron |
| Goalpara (*10*) | 2·7 | Diamond, graphite, α-iron, trolite, chromite |
| Novo Urei (*22*) | 1·9 | Diamond, graphite, α-iron, clinopyroxene |

The direct conversion of graphite to diamond is not a complex one. It is necessary only to increase the interatomic distance within individual carbon planes by 0.12 Å (Fig. 1) and to decrease

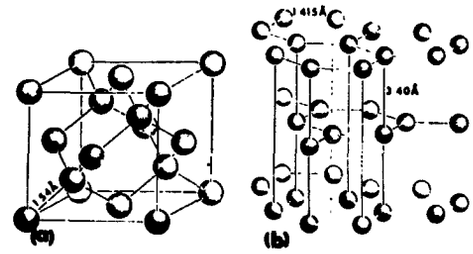


FIGURE 1.—Crystal structures of diamond (a) and graphite (b) [after Taylor (9)].

the interplanar spacing (*c*-axis) by about 1.86 Å. If a randomly oriented infinite distribution of graphite crystallites is assumed, it should be possible to determine whether a given diamond aggregate was produced by shock, inasmuch as only those graphite crystallites with *c*-axes parallel to the direction of motion of the shock will be converted to diamond (*6*). The resulting diamonds should exhibit a preferred crystallographic orientation. Diamonds formed by hydrostatic (gravitational) compression, on the other hand, will show no such structure since hydrostatic pressure acts isotropically in a medium. In fact, the situation may not be quite so simple because of

*Published in *Science*, 143 (3613): 1431-1434, March 27, 1964.

†Goddard Space Flight Center and Physikalisches Institut, Universität Bern, Sidlerstrasse 5, Bern, Switzerland.

complicating effects due to secondary and rarefaction (reflection) waves. Therefore, a preferred orientation should be apparent in at least some instances of diamonds formed by shock. The absence of such an orientation, however, cannot rule out the formation of diamonds by shock.

Specimens from all three ureilites were prepared for x-ray analysis by carefully carving or chipping them out so as not to affect the possible orientation of the crystallites within each fragment. The specimens, each of which weighed less than about 0.1 mg, were then examined by previously described diffraction techniques and equipment (*2*) to determine their composition. Fortunately the other phases in the specimens did not possess diffraction lines which would interfere with those of diamond.

After verification of the presence of diamond in the fragment, the specimen was studied with a Unicam model S.25 single crystal goniometer (*7*). In this experimental arrangement, a polycrystalline aggregate would be indicated by uniform ellipses centered on the x-ray outlet port, while preferred orientation would be evidenced by enhancement of segments of these ellipses. Large single crystals would be indicated by single spots, while deformed single crystals would be evidenced by asterism of these spots in directions approximately perpendicular to the ellipses.

Stationary x-ray photographs were taken every 10° with iron-filtered CoK$\alpha$ [for the (111) diamond planes] and zirconium-filtered MoK$\alpha$ [for the (220) and (311) planes of Goalpara diamond]. The angles $\rho$ and $\Phi$, corresponding to the ends of the preferred orientation zones in each set of normals to the reflecting planes, were measured (*8*) and then plotted on a polar stereographic net graduated in 2° increments (*9*).

The composition and sizes of the diamonds in the ureilites vary. All of the discrete diamond grains that I have analyzed are polycrystalline aggregates consisting of diamond, randomly oriented polycrystalline graphite, and other phases (Table 1).

The $\alpha^1$–$\alpha^2$ doublet of the diamond (331) plane provides a useful measure of the average crystallite size. From the lack of resolution of this doublet in Goalpara, Zachariasen (*10*) has estimated that the diamonds are about 100 Å. The diamonds from Dyalpur and Novo Urei on the other hand

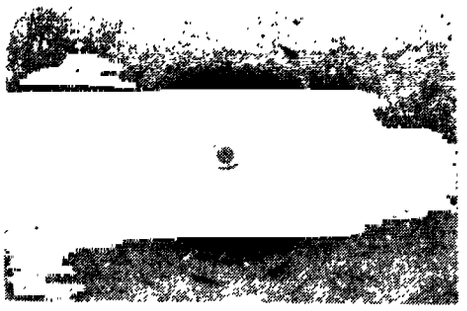

FIGURE 2.—Zirconium-filtered MoK$\alpha$ x-ray pattern of nonrotated diamond grain from Goalpara (×2). The graphite [002] plane shows that the graphite is polycrystalline and randomly oriented. The diamond [111], [220], [311], [400], and [331], on the other hand, shows preferred orientation.

are relatively large; those from Dyalpur are very much larger than 300 Å.

The diffraction pattern of a Goalpara diamond is shown in Fig. 2. The diamond reflections indicate pronounced preferred orientation whereas the graphite is randomly oriented. Stereographic projections of the diamond [111], [220], and [311] directions (Fig. 3 *a*, *b*, and *c*) reveal very striking patterns. The angles listed on each projection are those which that simple cubic plane makes with its (311) plane while the roman numerals indicate the multiplicity of these angles (*11*). Agreement is quite good and indicates that the oriented planes in the diamond are probably those parallel to (311).

It is important to note that the preferred orientation of diamond, which is indicative of formation by an anisotropic process, is present in Goalpara and (as will be demonstrated) also in Novo Urei. Natural terrestrial diamonds have never been reported to contain preferred orientation and in an extensive monograph (*12*) on these structures in terrestrial rocks and minerals there is no mention of this "fabric" in diamond.

Previous x-ray studies of hydrostatically produced synthetic diamond (*13*) do not indicate preferred orientation. The two samples which I studied do not show such orientation.

Synthetic diamonds produced by the anisotropic application of temperature or pressure or both, that is by flash-heating (*14*) or by a 300-kbar shock (Fig. 4), have oriented structures. The stereographic projections of these data show a grouping at about $\rho = 70°$ for the former and a

single "node" (compare with Fig. 5) at about 40° for the latter. Diamonds produced by shock at pressures significantly lower than 300 kbar are very polycrystalline and thus do not have detectable orientation. Thus, in accord with theory, the only diamonds that exhibit preferred orientation were produced by some anisotropic process.

The x-ray pattern of a diamond from Novo Urei (Fig. 6) is similar to that of a diamond produced

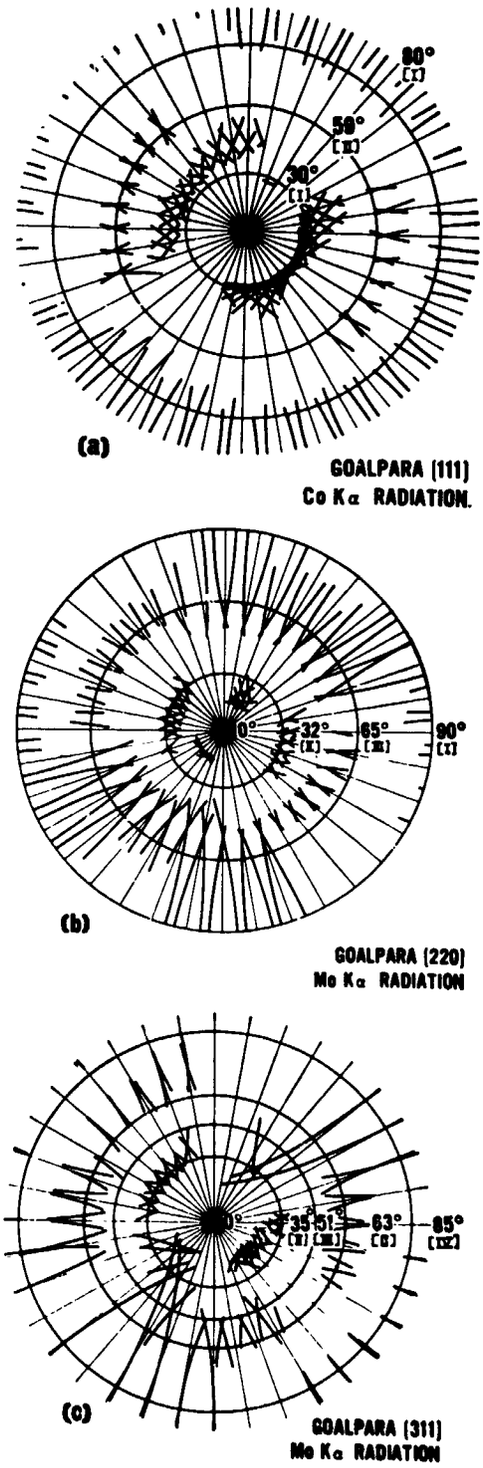


FIGURE 3.—Stereographic projections of Goalpara diamond (a) [111], (b) [220], and (c) [311]. The theoretical angles shown between each direction and the [311] are in agreement with the observed groupings.

FIGURE 4.—Iron filtered *CoKα* x-ray pattern of graphite partly converted to diamond by a 300-kbar shock. While the graphite is polycrystalline and randomly oriented, the diamond shows preferred orientation.

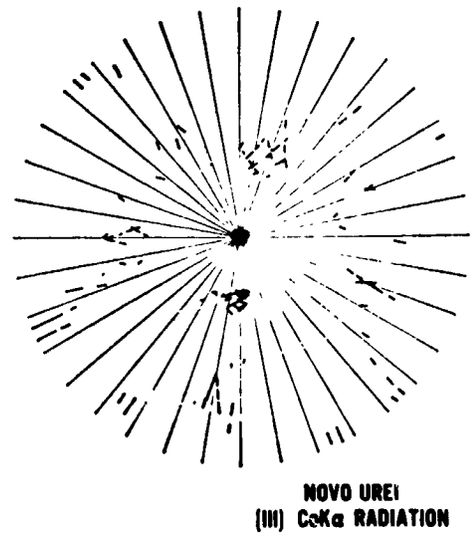


FIGURE 5.—Stereographic projection of Novo Urei diamond [111] directions. Although a pattern is evident, it is somewhat different in nature from that in Figure 3a.

FIGURE 6.—X-ray pattern (iron-filtered *CoKα*) of Novo Urei diamond. The diamond crystalline orientation is similar to that shown in Figure 4.

by a 300-kbar shock (Fig. 4). The stereographic projection of the Novo Urei data (Fig. 5) is different in character from that of Goalpara (Fig. 3*a*), possibly arising from a different alignment of the Novo Urei diamond relative to the x-ray beam or from actual differences in their pressure history.

Other phases in these two meteorites can yield information on the ureilites' pressure-temperature history. Olivine grains from Novo Urei (Fig. 7*a*) still appear to be reasonably good single crystals, although fractured to some extent. Goalpara olivine, on the other hand, is heavily fractured (Fig. 7*b*) and in some regions shows signs of recrystallization along grain boundaries, indicating excursion into temperatures well above its melting point. Since the only known natural process that can give rise to significant short-term anisotropic pressure is shock, it appears reasonable that the diamonds in at least these two ureilites were formed in this manner. It does not seem too extreme to suppose that the diamond formation and recrystallization of the olivine in Goalpara were contemporaneous. The data of Hughes and McQueen (*15*) on shocked olivine (dunite) indicate that even at the maximum pressure studied (700 kbar), the temperature reached only 720°. At pressures of 400 kbar, which is well in the diamond-forming region, the temperature rises to less than 200°. Thus, Goalpara, which was shocked more severely than Novo Urei, was probably exposed to pressures considerably in excess of 700 kbar.

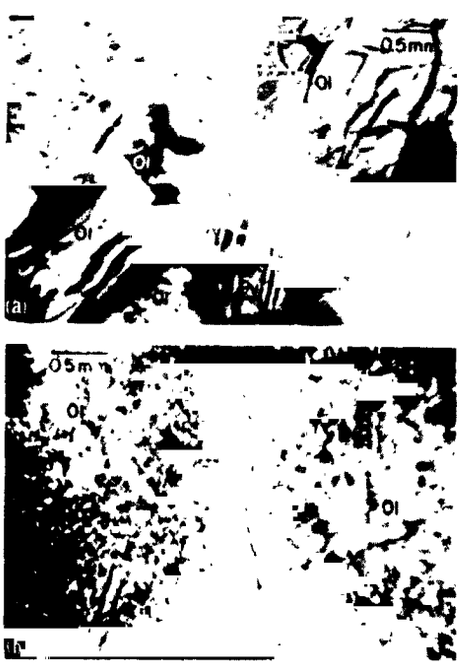


FIGURE 7.—Thin section of (a) Novo Urei ureilite and (b) Goalpara ureilite (both × 50). In (a), olivine (*Ol*) apparently still exists as single crystals although somewhat fractured. The clinopyroxene (*Py*) grains show cleavages. The opaque material is a graphite-diamond-kamacite mixture. In (b), the clinopyroxene grains are still apparently single crystals although some show an undulatory character under polarized light. Olivine grains (*Ol*) are badly shattered and in some locations are apparently recrystallized.

Diamond from Dyalpur shows no preferred orientation. One could suggest that the diamonds in this meteorite were formed by gravitational pressure rather than by shock as in Goalpara and Novo Urei. However, Dyalpur's textural similarity to the other two ureilites (particularly to Novo Urei (*16*) would argue against different modes of diamond formation by Occam's principle (see *17*).

Figure 3 (*a*, *b*, and *c*) shows certain features which may have bearing on the details of shock formation of diamond in Goalpara. First, some degree of symmetry seems evident in the distribution of groupings in each set of normals. Second, there are two "nodes" in the [220] that occur at about 20°. The former effect may be explained by lattice distortion of the diamond crystallites while the latter may be due to the presence of some diamonds formed by rarefaction waves or by secondary shocks. Another feature which may be of significance is that there appear to be only a few groupings at about 51°. These angles occur three times within the crystal (multiplicity III) and should be more prominent than the 63° angle which occurs twice (multiplicity II). This effect may only be an apparent one inasmuch as the zones in the Goalpara [311] are quite extended and markedly overlap the 51° line. The only other reasonable match for the groupings would be by planes parallel to the (220). Were these the oriented planes, however, we would not expect the group at about 30° in the [220] or at about 60° in the [111]. It is possible that the observed distribution of axes is due to some combination of oriented (110) and (311) planes, although the available evidence favors the interpretation of the oriented planes being parallel to (311) only.

From an examination of meteoritic and artificially shock-produced diamonds some sugges-

tions can be made concerning the mechanism of conversion of graphite to diamond by shock. At pressures well below 300 kbar, polycrystalline fine-grained diamond is apparently formed from polycrystalline graphite, possibly by the compression of rhombohedral graphite only (*4*). From pressures of about 300 kbar to well above 700 kbar the solid-state reaction involves conversion of the basal (001) plane of graphite to diamond (311), or possibly some combination of (311) and (110). Only at very high pressures may the conversion involve the hypothetical "metallic carbon" postulated by Libby (*18*) or possibly formation from shock-melted graphite.

The simultaneous presence of graphite, diamond, and kamacite in all of the ureilitic "diamonds" that I have investigated shows that these grains are not equilibrium assemblages. This raises the question of whether they represented an arrested stage in the conversion of graphite to diamond, or vice versa. Studies at 0 kbar (*19*) and 100 kbar (*20*) show that graphite formed from diamond shows preferred orientation. In at least two meteorites, however, it is the diamonds which are oriented, and in all diamantiferous meteorites the graphite is polycrystalline and randomly oriented. Thus, the graphite cannot have been formed from diamond. On the other hand, the theory predicts that, under favorable circumstances, shock-formed diamond would show preferential orientation. It thus appears that all meteoritic diamonds were formed by shock rather than by gravitational compression of graphite.

## REFERENCE AND NOTES


1. M. E. Lipschutz and E. Anders. *Geochim. Cosmochim. Acta* **24**, 83 (1961).
2. ———, *Science* **134**, 2095 (1961).
3. E. C. T. Chao, E. M. Shoemaker, B. M. Madsen, *ibid.* **132**, 220 (1960); E. C. T. Chao, J. Fahey, J. Littler, D. Milton, *J. Geophys. Res.* **67**, 419 (1962).
4. P. DeCarli and J. Jamieson, *Science* **133**, 1821 (1961).
5. Major minerals in these meteorites are olivine, $(Mg, Fe)_2SiO_4$, and clinopyroxene, (Ca. Mg, Fe) $SiO_3$. The meteorites contain, in addition, minor amounts of diamond, graphite, kamacite ($\alpha$-iron), troilite (FeS), and chromite ($FeCr_2O_4$). No other stony meteorites are known to contain diamond although well over 260 others have been examined for diamond by various investigators.
6. E. Anders, *Astrophys. J.* **134**, 1006 (1961).
7. Unicam Instruments Ltd., Cambridge, England.
8. N. F. M. Henry, H. Lipson, W. A. Wooster, *The Interpretation of X-Ray Diffraction Photographs* (Macmillan, New York, ed. 2, 1961), Fig. 108.
9. A. Taylor, *X-Ray Metallography* (Wiley, New York, 1961), Fig. 258.
10. H. C. Urey, A. Mele, T. Mayeda, *Geochim. Cosmochim. Acta* **13**, 1 (1957).
11. *International Tables for X-Ray Crystallography* (Kynock, Birmingham, England, 1959), vol. 2.
12. H. W. Fairbairn, *Structural Petrology of Deformed Rocks* (Addison-Wesley, Cambridge, Mass., ed. 2, 1949).
13. H. L. Pugh, J. Lees, J. A. Bland, *Nature* **191**, 865 (1961); J. A. Kohn and D. W. Eckart, *Am. Mineralogist* **47**, 1422 (1962); H. O. A. Meyer and H. J. Milledge, *Nature* **199**, 167 (1963).
14. F. P. Bundy, *J. Chem. Phys.* **38**, 631 (1963).
15. D. S. Hughes and R. G. McQueen, *Trans. Am. Geophys. Union* **39**, 959 (1958).
16. L. G. Kwascha, *Chem. Erde* **19**, 249 (1958).
17. Since this paper was submitted, A. P. Vinogradov and G. P. Vdovykin (*Geokhimiya* No. 8, 715 (1963)) independently confirmed the presence of diamond in the three ureilites. In addition, they have announced the discovery of diamond in yet another stony meteorite, Ghubara. When samples of this meteorite can be obtained, the properties of Ghubara's diamonds and olivine can be compared with those of the ureilites.
18. W. F. Libby, *Proc. Natl. Acad. Sci. U.S.* **48**, 1475 (1962).
19. H. J. Grenville-Wells, *Mineral. Mag.* **29**, 803 (1952).
20. V. M. Titova and S. I. Futergendler, *Soviet Phys. Cryst.* (*English Transl.*) **7**, 749 (1963).
21. M. E. Lipschutz, *Science* **138**, 1266 (1962).
22. M. Yerofeyev and P. Lachinov, *Compt. Rend.* **196**, 1679 (1888).


23. I am very grateful to Edward Anders for his many suggestions and comments during the course of this investigation. I also thank John Jamieson and Paul S. DeCarli, for many stimulating discussions, and E. C. T. Chao for his comments on the thin sections of the meteorites. I gratefully acknowledge the loan of the Dyalpur meteorite (E. Olsen, Chicago Natural History Museum) and the Goalpara and Novo Urei meteorites and thin sections (E. P. Henderson, U.S. National Museum). Gifts of synthetic diamonds produced by various techniques were received from R. H. Wentorf, Jr. (General Electric Research Laboratory), M. Waxman (National Bureau of Standards), and P. S. DeCarli (Stanford Research Institute). Miss Vida Milanovic and Miss Donna Reichel provided valuable assistance. The Materials Research and Development Section, NASA, lent equipment during the course of this research, which was supported in part by National Science Foundation grant G 14298.

N69-11724

# THE BACKWARD RECURRENCE METHOD FOR COMPUTING THE REGULAR BESSEL FUNCTION*

THOMAS E. MICHELS

The Backward Recurrence Method (suggested by Dr. J. C. P. Miller) for computing the regular Bessel function is investigated for real, imaginary, and complex arguments. From this, one can obtain the functions $J_n(z)$, $I_n(z)$, $ber_n(z)$, and $bei_n(z)$. A section has also been devoted to show the behavior of the functions for all arguments, and graphs are presented which show the magnitudes of the functions for all orders and real arguments between 0 and 6000.

## INTRODUCTION

Bessel's Equation is a second order differential equation with a singular point at $z=0$ and is of the form

$$\frac{d^2y}{dz^2} + \frac{1}{z}\frac{dy}{dz} + \left(1 - \frac{n^2}{z^2}\right)y = 0 .$$

Its two independent solutions are referred to as the regular and irregular Bessel functions (see Appendix A). Computation of the irregular function, $Y_n(z)$, is fairly easy, since one can use a straightforward recurrence relationship once he has two starting values (Reference 1) and obtain good accuracy for all n and z values. However, this method cannot be used to compute the regular function, $J_n(z)$, due to error accumulation in the recurrence. For large order and argument, asymptotic formulas prove to be too cumbersome; and in computer computation they are also inaccurate. But a method was devised by Dr. J. C. P. Miller (Reference 2) whereby an ingenious use of the recurrence relationship yields good values for all orders and arguments which fall within the limits of any particular computer.

It is the purpose of this report to explain the method and the need for it, as well as give details on the behavior and use of the recurrence formula and the accuracies obtained. (Hopefully it will also give insight into the behavior of the functions themselves.) This has been done for the argument z real, imaginary, and complex. In order to keep to this primary purpose, not much will be said about the irregular function. In general, however, the irregular functions behave just oppositely to the regular function in the range where $J_n(z)$ decreases or increases exponentially, while they behave similarly in the rest of the range. That is, if the magnitude of $J_n(z)$ *decreases exponentially*, the irregular function *increases;* and when $J_n(z)$ oscillates, the irregular function does the same.

It should be noted that the recurrence formula computes a sequence of Bessel functions as a function of n, the order, rather than the usual tabulations which are a function of z. Throughout this discussion, where the order is referred to as n in $J_n(z)$, it should be taken to mean any value of n (integer, half orders, one-third orders, etc.), unless otherwise specified. Also m is always greater than n, and $z(=x+iy)$ is regarded as a complex number. If just the real numbers are being referred to, x will be used, and for the imaginary iy will be used.

Reference is made from time to time to single and double precision. All tests were made on an IBM 7090 computer which has a word length of 36 binary bits. In the floating point mode, single precision arithmetic is done with 9 decimal digits and double precision with 18 decimal digits. This is implied throughout.

## EXPLANATION OF THE METHOD FOR COMPUTING THE REGULAR BESSEL FUNCTION

### Importance of the Backward Recurrence

It is well known that both the regular $J_n(z)$ and the irregular $Y_n(z)$, Bessel functions (see Appen-

*Published in *NASA Technical Note* D 2141, March 1964.

dix A) obey the recurrence relationship

$$F_{n+1}(z) + F_{n-1}(z) = \frac{2n}{z} F_n(z) .$$

If the recurrence is started with either $J_0(z)$, $J_1(z)$ or $Y_0(z)$, $Y_1(z)$, in the computer where there are a fixed number of digits, good accuracy is obtained from $n \leq z$ for both functions. This accuracy continues for the irregular function since $Y_n(z)$ *increases exponentially* for $n > z$, and the number of accurate digits is maintained or increased; but for the regular function at this point accuracy is lost very rapidly, since $J_n(z)$ *decreases* exponentially for $n > z$.

Actually, throughout the recurrence, by recurring in either direction, we obtain numbers which represent solutions to Bessel's equation, but they are linear combinations of the regular and irregular solution. Therefore, the error introduced in recurring for the regular function is actually some constant times the irregular function.

In using the formula for recurring Bessel functions, and recurring in the direction of increasing n, the number of digits of accuracy is lost very rapidly. This is not even taking into account loss of accuracy due to round off.

Let us take an extreme case and look at it this way: Assume we are in a computer which produces 8 significant digits and we are using the recurrence formula below with a small error in each term on the right, say in the last decimal place. We ask what the error $\epsilon_{n+1}$ is in our new term $F_{n+1}$:

$$F_{n+1}(z) + \epsilon_{n+1} = \frac{2n}{z} \left[F_n(z) + \epsilon_n\right] - \left[F_{n-1}(z) + \epsilon_{n-1}\right] .$$

Let us assume reasonable values—say for the second term on the right $F_{n-1}(z) \approx 10^{-1}$—and make the error the best we could have, and therefore in the last significant digit, or $\epsilon_{n-1} \approx 10^{-8}$. Let the magnitude of $F_n(z)$ be smaller, say $10^{-2}$, again with the error being in the last place, or $\epsilon_n \approx 10^{-9}$. Now we recur to a smaller number $F_{n+1}(z)$, say of the magnitude $10^{-3}$. Still, there was an error at $10^{-8}$, and therefore $\epsilon_{n+1} \approx 10^{-8}$. The $(n-1)$ term had 7 accurate digits and two terms away, at $(n+1)$, the number of accurate digits is reduced by two and our error is now in the sixth significant digit.

A similar argument can be made to show that when we recur to numbers of increasing magnitude, the number of significant digits of accuracy is actually increased, or at least stays equal to the maximum number of digits available in the computer. Therefore, the use of the foregoing recurrence formula to compute the regular Bessel functions in the region where they decrease in magnitude is not acceptable as it stands.

### Statement of Method

#### *Miller's Method*

A scheme was devised by Miller (Reference 2) whereby the recurrence formula could be used and accuracy still maintained throughout the whole range of n values. We start at some m, with

$$F_{m+1}(z) = 0 ,$$

$$F_m(z) = a ,$$

where a is any constant, and use the recurrence formula, but in *decreasing* magnitude in n (where we are recurring to larger numbers and therefore increasing accuracy) and generate a series of functions $F_n(z)$, $F_{n-1}(z), \cdots F_0(z)$ at some $n < m$, by

$$F_{n-1}(z) = \frac{2n}{z} F_n(z) - F_{n+1}(z)$$

which are all a constant multiple of the regular Bessel function. That is,

$$F_n(z) = \alpha J_n(z) .$$

#### *Proof*

Since the recurrence formula follows for both the regular and irregular function, we can say that it follows for a linear combination of them also, or

$$F_n(z) = \alpha J_n(z) + \beta Y_n(z) .$$

Since

$$F_{m+1}(z) = 0 = \alpha J_{m+1}(z) + \beta Y_{m+1}(z) .$$

we have

$$F_n(z) = \alpha \left[ J_n(z) - J_{m+1}(z) \frac{Y_n(z)}{Y_{m+1}(z)} \right] .$$

Therefore at some $n < m$

$$J_{m+1}(z) \frac{Y_n(z)}{Y_{m+1}(z)} \approx 0$$

and hence

$$F_n(z) = \alpha J_n(z)$$

to any desired degree of accuracy. Therefore all that remains to obtain the regular Bessel function is to determine $\alpha$.

### *Determination of $\alpha$*

There are a number of ways in which $\alpha$ can be determined. One method is to use an addition formula such as the following (Reference 3):

$$2^{\nu} \sum_{k=0}^{\infty} (\nu + 2k) J_{\nu+2k}(z) z^{-\nu} \frac{\Gamma(\nu + k)}{k!} = 1 .$$

which, of course, for integral orders or $\nu = 0$, becomes

$$J_0(z) + 2 \sum_{k=1}^{\infty} J_{2k}(z) = 1 .$$

or, for $F_k(z) = \alpha J_k(z)$,

$$F_0(z) + 2 \sum_{k=1}^{\infty} F_{2k}(z) = \alpha .$$

The summation can be carried out to $k = n$ and will give $\alpha$ the particular accuracy desired. Here n implies the number of the term $< m$ where $F_n(z) = \alpha J_n = \alpha J_n(z)$ (see page 1251).

Of course, if an expression is known for a particular $J_n(z)$ value, say

$$J_{1/2}(z) = \left(\frac{2}{\pi z}\right)^{1/2} \sin z .$$

then $\alpha$ is more easily determined by

$$\alpha = \frac{F_{1/2}(z)}{J_{1/2}(z)} .$$

For z real, the foregoing addition formula proves to be fairly accurate in the determination of $\alpha$; however, for z complex or imaginary, using the addition formula in the comput ' proves to be inaccurate (see page 1252).

The value of the Bessel function is only as accurate as $\alpha$, since each term is divided by $\alpha$; therefore the computation of $\alpha$ is extremely critical.

### *Error Introduced from Additional Formulas for Complex or Imaginary Argument*

In looking at the regular Bessel function of complex or imaginary argument, we observe that

$$|J_0(z)| \approx 0.1 e^{y}$$

and

$$\lim_{n \to \infty} J_n(z) = 0 .$$

As n increases, $J_n(x+iy)$ decreases at an exponential rate almost immediately for $|z| < 10$, and for larger z it takes a few more terms before the exponential decrease. It is almost a logarithmic spiraling into the origin of the complex plane.

With this in mind, assume we are performing the addition formula for integral orders. We want to investigate the error that will exist in the determination of $\alpha$, assuming an error in some $F_n(z)$.

In our downward recurrence we have the following:

$$F_0(z) + \epsilon_0 + 2 \sum_{n=1}^{\infty} \left( F_{2n}(z) + \epsilon_{2n} \right) = (1 + \epsilon) \alpha .$$

Assume the error to be in the last digit in all $F_n(z)$, which of course is the best we could expect in the

computer. Upon factoring out $\alpha$, we have

$$\alpha(J_0(z)+\epsilon_0) + 2\alpha\sum_{n=1}^{\infty}(J_{2n}(z)+\epsilon_{2n}) = \alpha(1+\epsilon) ;$$

and since

$$J_0(z) + 2\sum_{n=1}^{\infty} J_{2n}(z) = 1 ,$$

we have

$$\epsilon_0 + 2\sum_{n=1}^{\infty}\epsilon_{2n} = \epsilon ,$$

where $\epsilon$ is now the error we introduce in computing the number 1.

Since the errors decrease at an exponential rate as n increases,

$$\epsilon \approx \epsilon_0 ,$$

which says the error we introduce in computing the digit 1 is of the order of the error in $J_0(z)$. Therefore $\alpha$ will be in error by this factor. It can be seen that if $J_0(z)$ had five figure accuracy and was of the order of magnitude $10^7$, the summation would actually become meaningless.

Thus, to determine $\alpha$ one of the Bessel function values must be computed. This can be done by an asymptotic formula which simplifies for $J_0(z)$, or the following integral relationship can be used:

$$J_n(z) = \frac{1}{\pi}\int_0^{\pi}\cos(n\theta - z\sin\theta)\,d\theta ,$$

or

$$J_0(z) = \frac{1}{\pi}\int_0^{\pi}\cos(z\sin\theta)\,d\theta .$$

Integration of the expression for $J_0(z)$ by the trapezoidal rule generally yields sufficient accuracy, especially for large z values for which the slopes of the functions become large. For larger z values, of course, $\Delta\theta$ must be decreased. Integration performed by using the trapezoidal rule on various $|z|$ values (less than 75) and with $\Delta\theta=\pi/1000$ gave accuracy in the seventh and eighth digit.

When using the recurrence formula for $I_n(y)$

$$G_{n-1}(y) - G_{n+1}(y) = \frac{2n}{y}G_n(y) ,$$

the summation

$$e^y = I_0(y) + 2\sum_{n=1}^{\infty} I_n(y)$$

can be used to determine $\alpha$. This formula is of course used in the same manner as are the $F_n(z)$ functions. That is,

$$G_{m+1}(y) = 0 ,$$

$$G_m(y) = a ,$$

and, at the same n,

$$G_n(y) = aI_n(y)$$

to any desired degree of accuracy, and

$$ae^y = G_0(y) + \sum_{n=1}^{\infty} G_n(y).$$

No tests were made with this recurrence formula.

### Determination of n versus m Values

We cannot give a definite value to the number of terms necessary in the recurrence to obtain good values for $J_n(z)$. It depends, of course, on z as well as the value given to a in $F_m(z)=a$, and also upon the particular accuracy desired.

The following empirical formulas will give values for m which generate a sequence $J_0(z),\cdots,J_n(z)$, where $J_n(z)$ is of the magnitude $10^{-21}$, that have the accuracies described in a later section. This is assuming computation is done in single

precision and $a = 10^{-20}$.

$$m = 5.0\,|z| + 15 \quad \text{for} \quad 0.1 \le |z| < 10 ; \tag{1}$$

$$m = 1.48\,|z| + 48 \quad \text{for} \quad 10 \le |z| < 150 ; \tag{2}$$

$$m = 1.05\,|z| + 112 \quad \text{for} \quad 150 \le |z| \le 6000 . \tag{3}$$

It is felt that Equation 3 can be extended indefinitely for all values of $z \ge 150$; however, tests did not go beyond the range given.

For all values of z and n the following general rule can be employed: *For single precision accuracy, a safe* m *vs* n *is obtained when the following condition is met:*

$$\frac{J_m(z)}{J_n(z)} < 10^{-5} .$$

Table 1 gives various n vs m values for real arguments in the range from 0.01 to 6000.

## Details on Accuracy

### z Real

It has been said that accuracy is gained in the downward recurrence until the point where oscillations begin to occur. If the correct n vs m has been used for a particular x, then at this n value $F_n(x) = \alpha J_n(x)$ to any desired degree of accuracy. From here down to $J_0(x)$, accuracy loss is due only to round off and truncation in the computer. But, in this region, the functions oscillate usually between +0.3 and the number of accurate digits

Table 1

The integral order n of the regular Bessel function having a particular magnitude in the range where $n > x$ is given for various x values, and the recurrence term m which will yield eight place accuracy for all orders from zero to n. Also tabulated is the value of m which yields overflow in the computer for the downward recurrence at $F_0(x)$. These values were obtained using double precision with $F_m(x) = 10^{-20}$.

| Argument x | Magnitude of $J_n(x)$, where $n > x$, for particular orders n and recurrence terms m | | | | | | | | | | m value yielding overflow |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $10^{-2}$ | | $10^{-8}$ | | $10^{-16}$ | | $10^{-24}$ | | $10^{-32}$ | | |
| | n | m | n | m | n | m | n | m | n | m | |
| .01 | 1 | 4 | 4 | 6 | 6 | 8 | 9 | 11 | 11 | 13 | 19 |
| .05 | 1 | 5 | 5 | 7 | 8 | 10 | 11 | 13 | 14 | 16 | 22 |
| .10 | 2 | 5 | 5 | 7 | 9 | 11 | 12 | 14 | 16 | 18 | 26 |
| .50 | 3 | 7 | 8 | 11 | 13 | 15 | 17 | 19 | 21 | 23 | 34 |
| 2.0 | 6 | 12 | 12 | 16 | 19 | 22 | 25 | 28 | 30 | 33 | 47 |
| 5.0 | 9 | 13 | 18 | 23 | 26 | 30 | 34 | 38 | 40 | 43 | 60 |
| 8.0 | 14 | 16 | 22 | 28 | 32 | 37 | 41 | 45 | 48 | 51 | 70 |
| 12.0 | 17 | 20 | 28 | 34 | 39 | 44 | 48 | 53 | 57 | 61 | 81 |
| 20.0 | 25 | 29 | 39 | 46 | 51 | 57 | 62 | 67 | 72 | 76 | 100 |
| 40.0 | 46 | 55 | 63 | 72 | 79 | 86 | 92 | 98 | 104 | 109 | 138 |
| 100.0 | 109 | 119 | 130 | 143 | 152 | 163 | 170 | 179 | 186 | 193 | 229 |
| 300.0 | 310 | 321 | 334 | 361 | 374 | 389 | 398 | 411 | 420 | 431 | 485 |
| 500.0 | 511 | 530 | 555 | 583 | 587 | 604 | 615 | 631 | 642 | 655 | 715 |
| 1000. | 1012 | 1047 | 1063 | 1088 | 1107 | 1128 | 1145 | 1164 | 1177 | 1192 | 1270 |
| 2000. | 2015 | 2055 | 2080 | 2115 | 2135 | 2165 | 2185 | 2210 | 2225 | 2245 | 2340 |
| 3000. | 3015 | 3065 | 3090 | 3127 | 3154 | 3208 | 3208 | 3235 | 3255 | 3277 | 3385 |
| 4000. | 4015 | 4070 | 4095 | 4140 | 4170 | 4205 | 4230 | 4260 | 4280 | 4305 | 4425 |
| 6000. | 6015 | 6080 | 6110 | 6160 | 6195 | 6235 | 6260 | 6295 | 6320 | 6350 | 6485 |

becomes more or less as the recurrence goes from larger to smaller numbers. Therefore, accuracy loss due to round off is not quite as bad as it might seem. In fact, when recurring for $x=6000$, and $m=6400$, the error only crept over to the fifth significant digit; this was using single precision in the computer.

Particular values were checked in already existing tables (References 3, 4) for $0.05 \leq x \leq 100$ and $0 \leq n \leq 100$. Values having greater x and n were checked against values obtained using the recurrence in double precision. The accuracies obtained were the following:

| Argument | Accuracy |
|---|---|
| | within a few figures |
| $0.05 < x \leq 30$ | in the 8th digit |
| $30 < x \leq 100$ | at least 7 digits |
| $100 < x \leq 1000$ | at least 5 digits |
| $1000 < x \leq 6000$ | at least 4 digits |

These accuracies are the worst obtained over the whole range of orders out to where $J_n(x) \approx 10^{-38}$.

Using double precision, of course, would give approximately eight more digits of accuracy for all arguments tabulated. It seemed unnecessary to check the recurrence for real arguments greater than 6000, in view of their sparse usage.

The accuracy stated above is obtained only if $\alpha$ in

$$F_n(x) = \alpha J_n(x)$$

is computed to at least the number of digits stated. Of course, if $\alpha$ is only good to a few digits then the Bessel function can be no better.

### *z Complex or Imaginary*

In the downward recurrence for imaginary and complex argument $(x+iy)$, where $|y/x|$ is not small, the magnitudes of the numbers increase over the whole range of n values. Therefore, accuracy is gained or maintained over the whole range. These functions were checked in tables of Thomson functions for ber(x) and bei(x) and in formulas for $J_{1/2}(z)$, $J_{3/2}(z)$, $J_{5/2}(z)$. These all proved to be accurate to seven digits for various $|z|$ values ranging from .05 to 100.

For argument $(x+iy)$ where $|y/x|$ is small, approximately, less than 0.1) the function behaves similarly to the real function in that the magnitudes oscillate and then continually decrease when n becomes greater than z. In general the accuracies should behave similarly to the accuracies of the real function.

### Details on Use of the Recurrence
### *Choosing the Value for $\alpha$*

When using the above method for computing the regular Bessel function, the value we give to a is completely arbitrary. In other words the method "works" for any value of a. However, as a general rule, we should use as small a value, in magnitude, as is practical for the computer. This enables us to start out farther in the downward recurrence or use a larger m value and thereby obtaining higher order Bessel functions—and, of course, with less chance of overflow in recurring to numbers which are too large for the particular computer.

Consider a sequence of functions, say for real z,

$$J_0(x),\ J_1(x) \cdots J_n(x),\ J_{n+1}(x)\ .$$

The magnitudes of the functions may vary from approximately 0.1 for $J_0(x)$ down to say $10^{-38}$ for $J_n(x)$. That is, the values cover a range equal to approximately $10^{38}$. Since the same range would be covered in our downward recurrence, the magnitude of $F_0(x)$ depends completely on the value of a at any particular m. If $a=10^{-12}$, then $F_0(x)$ would be 38 orders of magnitude away or $10^{25}$, and so on. Therefore, to compute larger order Bessel functions, a smaller value for

$$F_m(x) = a\ ,$$

should be used. It should be noted that the value given to a will not increase or decrease accuracy.

### *z Real*

As has been stated, $J_0(x)$ does not become greater than 1 for all x, and

$$\lim_{n \to \infty} J_n(x) = 0\ ,$$

the values oscillate usually between $\pm 0.3$ until n becomes greater than x, then they stay positive

and begin to decrease at exponential rate to zero at $n = \infty$. The recurrence must start beyond this point for the method to succeed.

As a general rule, for all x, recurrence can begin at an m value for a particular order n where

$$\frac{F_m(x)}{F_n(x)} = \frac{J_m(x)}{J_n(x)} < 10^{-s} ,$$

and still obtain eight figure accuracy for $J_n(x)$.

### *z Complex*

In contrast to the function of real argument,

$$|J_0(z)| \approx 0.1\, e^{y} ,$$

but still

$$\lim_{n\to\infty} J_n(x) = 0 ,$$

and the magnitude of the complex function decreases continually, slowly at first for low orders

Table 2

Values obtained for $J_n(10)$ from the forward recurrence in double and single precision and the backward recurrence in double precision.

| n | $J_n(10)$ | | |
|---|---|---|---|
| | Forward Recurrence Single Precision | Forward Recurrence Double Precision | Backward Recurrence Double Precision |
| 0 | -0.24593576E-00 | -0.24593576E-00 | -0.24593576E-00 |
| 1 | 0.43472745E-01 | 0.43472745E-01 | 0.43472745E-01 |
| 2 | 0.25463031E-00 | 0.25463031E-00 | 0.25463031E-00 |
| 3 | 0.58379377E-01 | 0.58379379E-01 | 0.58379379E-01 |
| 4 | -0.21960268E-00 | -0.21960268E-00 | -0.21960268E-00 |
| 5 | -0.23406152E-00 | -0.23406152E-00 | -0.23406152E-00 |
| 6 | -0.14458837E-01 | -0.14458842E-01 | -0.14458842E-01 |
| 7 | 0.21671092E-00 | 0.21671092E-00 | 0.21671092E-00 |
| 8 | 0.31785411E-00 | 0.31785412E-00 | 0.31785412E-00 |
| 9 | 0.29185566E-00 | 0.29185568E-00 | 0.29185568E-00 |
| 10 | 0.20748607E-00 | 0.20748610E-00 | 0.20748610E-00 |
| 11 | 0.12311649E-00 | 0.12311652E-00 | 0.12311652E-00 |
| 12 | 0.63370191E-01 | 0.63370255E-01 | 0.63370255E-01 |
| 13 | 0.28971968E-01 | 0.28972083E-01 | 0.28972083E-01 |
| 14 | 0.11956926E-01 | 0.11957163E-01 | 0.11957163E-01 |
| 15 | 0.45074252E-02 | 0.45079730E-02 | 0.45079730E-02 |
| 16 | 0.15653493E-02 | 0.15667561E-02 | 0.15667561E-02 |
| 17 | 0.50169241E-03 | 0.50564667E-03 | 0.50564667E-03 |
| 18 | 0.14040489E-03 | 0.15244248E-03 | 0.15244248E-03 |
| 19 | 0.37652134E-05 | 0.43146276E-04 | 0.43146277E-04 |
| 20 | -0.12609708E-03 | 0.11513367E-04 | 0.11513369E-04 |
| 21 | -0.50815354E-03 | 0.29071915E-05 | 0.29071994E-05 |
| 22 | -0.20081478E-02 | 0.69683732E-06 | 0.69686851E-06 |
| 23 | -0.83276965E-02 | 0.15889271E-06 | 0.15902198E-06 |
| 24 | -0.36299255E-01 | 0.34069179E-07 | 0.34632629E-07 |
| 25 | -0.16590872E-00 | 0.46393399E-08 | 0.72146349E-08 |
| 26 | -0.79324437E-00 | -0.10872479E-07 | 0.14405452E-08 |
| 27 | -0.39589619E 01 | -0.61176230E-07 | 0.27620052E-09 |
| 28 | -0.20585150E 02 | -0.31947916E-06 | 0.50937552E-10 |
| 29 | -0.11131787E 03 | -0.17279071E-05 | 0.90497669E-11 |
| 30 | -0.62505851E 03 | -0.97023620E-05 | 0.15510961E-11 |
| 31 | -0.36390331E 04 | -0.56486385E-04 | 0.25680948E-12 |
| 32 | -0.21936946E 05 | -0.34051320E-03 | 0.41122714E-13 |
| 33 | -0.13675742E 06 | -0.21227981E-02 | 0.63758926E-14 |
| 34 | -0.88066203E 06 | -0.13669954E-01 | 0.95817661E-15 |
| 35 | -0.58517443E 07 | -0.90832891E-01 | 0.13970838E-15 |
| 36 | -0.40081548E 08 | -0.62216029E 00 | 0.19782068E-16 |
| 37 | -0.28273539E 09 | -0.43887211E 01 | 0.27225057E-17 |
| 38 | -0.20521604E 10 | -0.31854376E 02 | 0.36447453E-18 |
| 39 | -0.15313683E 11 | -0.23770464E 03 | 0.47500695E-19 |
| 40 | -0.11739457E 12 | -0.18222410E 04 | 0.60308953E-20 |

until, approximately when n becomes greater than x/2, they decrease at an exponential rate. The role that x plays will be discussed in a later section. Therefore the downward recurrence can begin anywhere in the range of n values as long as the condition

$$\frac{F_m(z)}{F_n(z)} < 10^{-5} ,$$

is met to give eight figure accuracy in $J_n(z)$.

### z Imaginary

The function of imaginary argument behaves similarly to the function of complex argument in that they continually decrease in magnitude and, approximately at z/2, they decrease exponentially. For integral orders, the functions oscillate between being pure real and pure imaginary; but for orders that are not integral, $J_n(iy)$ possess both real and imaginary parts. The integral orders are usually multiplied by $e-in\,\pi/2$ to give the real function, $I_n(y)$, which is widely used. That is,

$$I_n(y) = e^{-in\pi/2} J_n(iy) .$$

More will be discussed on the function $I_n(y)$ in the next section. An example of the method will be found in Table 2, which gives $J_n(10)$ by the forward recurrence technique in both single and double precision, and by the backward recurrence technique in double precision.

## NUMERICAL BEHAVIOR OF THE REGULAR BESSEL FUNCTION

In seeking to present a complete picture of the regular solutions to Bessel's equation

$$y'' + \frac{1}{z} y' + \left(1 - \frac{n^2}{z^2}\right) y = 0$$

it seemed that a better understanding of the solutions was brought about when first they were observed as functions of n, and then as functions z. Then they could be brought together and a better picture could be visualized for all n and z.

### z Real

Two-dimensional graphs have been drawn (Appendix B) showing various lines of constant magnitude for all real z and n for $0<x\leq6000$. These can be extended to larger x, since the lines are fairly linear.

It n be seen from a plot of $J_n(x)$ vs. n (Figure 1) that the functions oscillate between $\pm1$, with

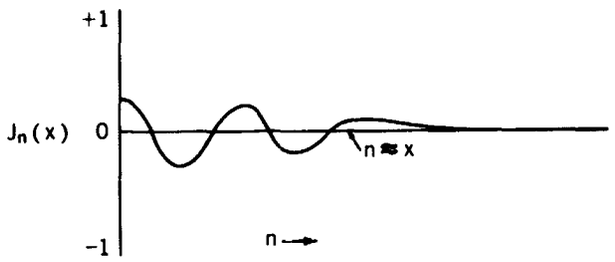


FIGURE 1.—Plot of $J_n(x)$ vs. n, n continuous.

the oscillations dampening in the range $0\leq n\leq x$. Then for $n\geq x$ the values stay positive and begin to decrease very rapidly (here is where forward recurrence begins to fail), until at $n=\infty$, $J_n(x)=0$. This is true for all values of x with the exception $J_n(0)=0$, and $J_0(0)=1$. Let us now consider the function more closely.

For $5<x<n$, $J_n(x)$ oscillates between $\pm0.3$ and for $0<x\leq5$, $n\leq3$, $J_n(x)$ oscillates between $\pm0.2$ and for all n, $J_n(n)\leq\cdot1$. It can be seen that dampening of the oscillations takes longer for larger x, since it takes more terms for n to become similar to x.

On the other hand, it can be seen from a plot of $J_n(x)$ vs. x (Figure 2) that the functions start

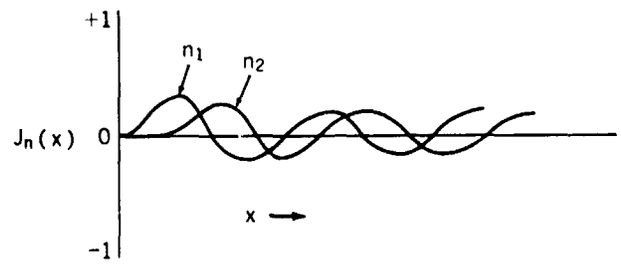


FIGURE 2.—Plot of $J_n(x)$ vs. x, $n_1<n_2$.

at $x=0$, (except that $J_0(0)=1$) and increase very slowly; the larger the n value the slower it increases, staying positive and never exceeding 0.1 until $n\approx x$. At this point they continue to increase, but as x increases oscillations begin. For $n>3$ and all x, the oscillations are between $\pm0.3$.

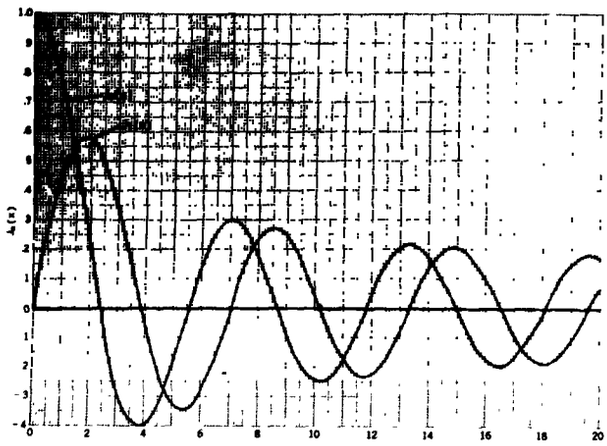

FIGURE 3.—Detailed plot of $J_0(x)$ and $J_1(x)$ vs. x, $0 \leq x \leq 20$.

Figure 3 shows a detailed plot of $J_0(x)$ and $J_1(x)$ versus x.

### z Complex

The magnitude of the function for complex argument $z = x+iy$ begins at

$$|J_0(x+iy)| \approx 0.1\, e^y$$

for any z and, where $|y/x|$ is not small (see page 9)

$$\lim_{n\to\infty} J_n(x+iy) = 0 .$$

The functions continually decrease in magnitude as n becomes larger and, at approximately where $n > |z|/2$, they decrease at an exponential rate.

If we plot the modulus of $J_n(z)$ vs. n, we get the exponential curve shown in Figure 4. The

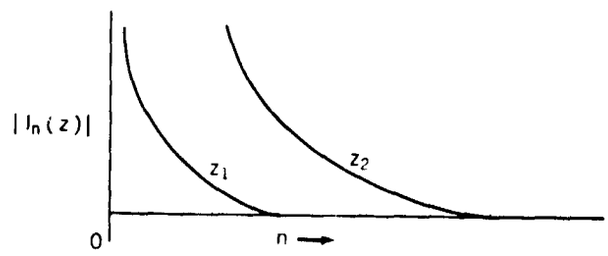


FIGURE 4.—Plot of $|J_n(x+iy)|$ vs. n.

effect of x in the argument is that for larger x, the gradient of the slope becomes less.

In considering a picture of the function of complex argument, it seems that the easiest and perhaps the best approach is set

$$J_n(x+iy) = u+iv$$

and plot the functions in both planes.

In Figure 5, the spiral is started at $J_0(x+iy)$, and as n increases the function spirals into the

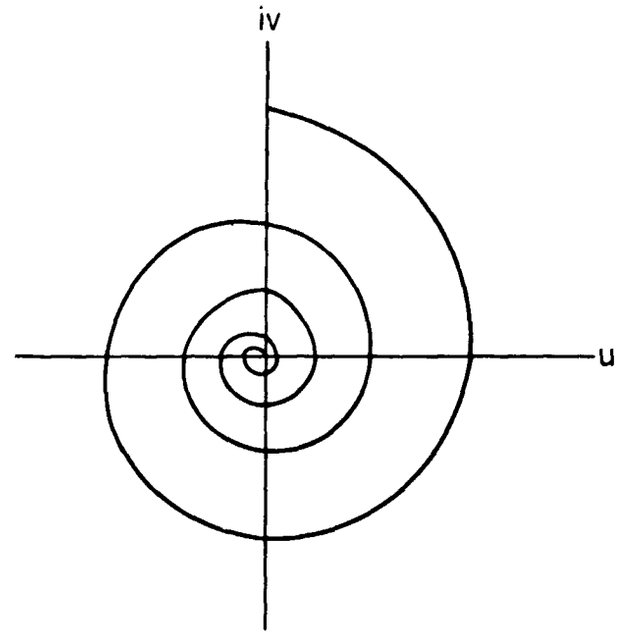


FIGURE 5.—Projection of $J_n(x+iy)$ on the u, iv plane as a function of n, n continuous.

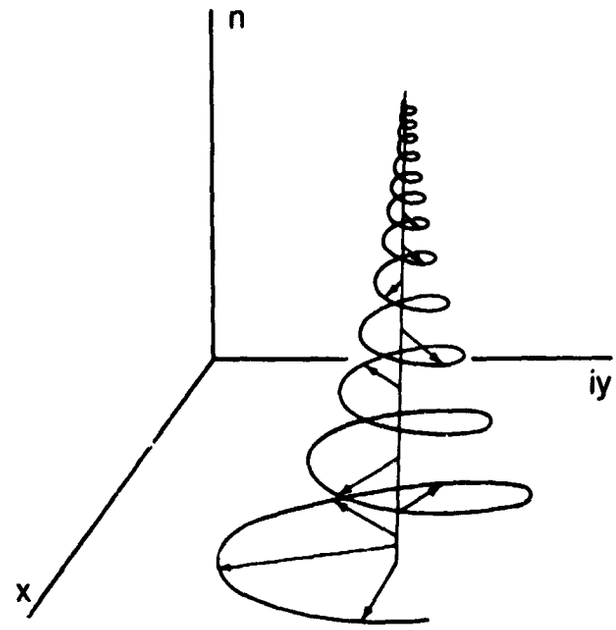


FIGURE 6.—Plot of $J_n(x+iy)$ in the x, iy, n space as function of n, n continuous.

origin as a limit. Figure 6 is again a plot of $J_n(x+iy)$ as a function of n but now it is the three dimensional picture in x, iy, and n.

If the spiral is moved around in the x, iy plane, the base of the spiral becomes larger as y increases, and becomes more spread out or flatter as x increases. If movement was directed along one of the 45 degree lines, the slope and the width would change at approximately the same rate. It should be noted that for movement along the 45 degree lines the plot would be of the Thomson Functions, $ber_n(z)$, $bei_n(z)$.

The direction of the spiraling is not significant, for it changes direction as the argument of z changes. The functions of arguments (x+iy) and (x−iy) would spiral into the origin in opposite directions.

If a three dimensional plot is made of $J_n(x+iy)$ in the x, iy plane, contour lines of constant magnitude would give the effect of a crater (Reference 5, page 127).

For argument (x+iy) where |y/x| is small, the modulus of $J_n(x+iy)$ oscillates until n becomes greater than z and at this point, it stays in the first quadrant of the u, iv plane, continually decreasing to the origin at $n = \infty$.

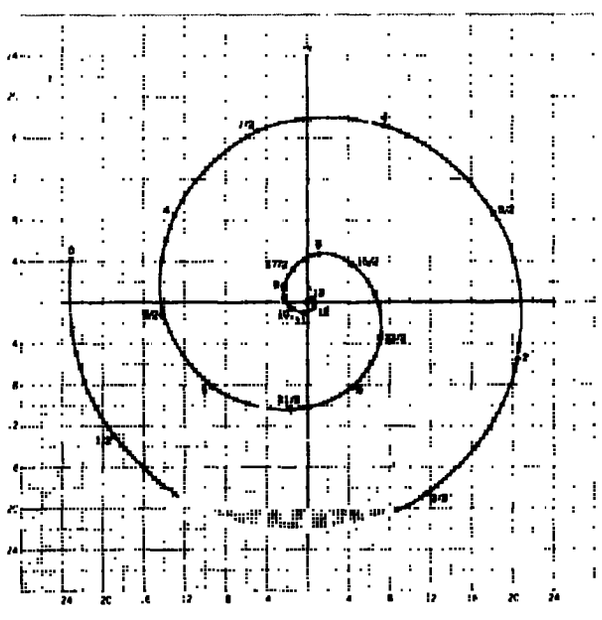

FIGURE 7.—Detailed plot of the projection of $J_n(10+i10)$ on the u,iv plane as a function of n.

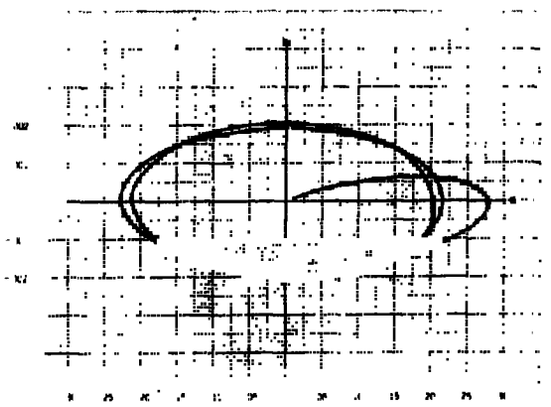

FIGURE 8.—Detailed plot of the projection of $J_n(15+6.01)$ on the u,iv plane vs. n.

### z Imaginary

The regular Bessel function of imaginary argument $J_n(iy)$ behaves similarly to the function of complex argument. $|J_0(iy)|$ is of the order of magnitude $0.1e^y$ with

$$\lim_{n\to\infty} J_n(iy) = 0 ,$$

and the function spirals in toward the origin of the u, iv plane. The figures for imaginary argument would be similar to the complex argument with the integral orders lying on the axis and half orders displaced by an angle $\pi/4$, etc. In the u, iv plane they would appear as in Figure 9. Here again the direction of the spiral is determined by the argument of z, or $\pm\pi/2$. These functions

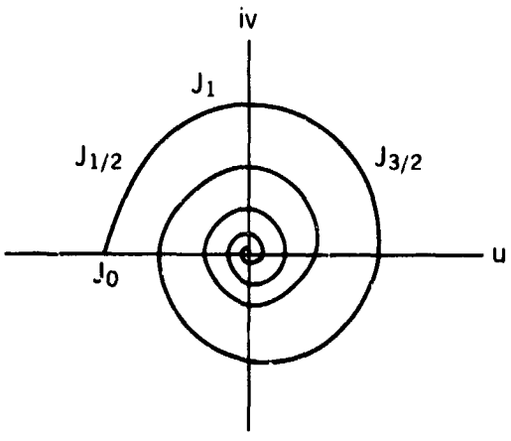


FIGURE 9.—Projection of $J_n(iy)$ on the u,iv plane.

are really solutions of the equation

$$y'' + \frac{1}{z} y' - \left(1 + \frac{n^2}{z^2}\right) y = 0 ,$$

which occurs frequently in problems of mathematical physics; however, it is usually desirable to express the solutions in real form.

The solutions $J_n(iy)$, $Y_n(iy)$ are not always real, as was shown for the regular function; however the function $e^{-in\pi/2}J_n(iy)$ is always real and is a solution of the equation. In effect the imaginary values are rotated thru an angle of $\pi/2$. These functions are the $I_n(y)$ functions and are defined by

$$I_n(y) = e^{-in\pi/2} J_n(iy) ;$$

they obey the recurrence formula

$$I_{n-1}(y) - I_{n+1}(y) = \frac{2n}{y} I_n(y)$$

the use of which was discussed in a previous section.

## CONCLUDING REMARKS

It has been shown that the backwa.d recurrence is an accurate and especially easy method for computing the regular Bessel Function. Indirectly, this report came about through a need for and investigation of computing the 1/2 order Bessel functions for the intensity functions of light scattering by spherical particles. Since the calculations required the Bessel functions varying in the order, the recurrence was especially suitable. The backward recurrence method should be a very useful tool to those requiring Bessel functions of any order or argument:

The magnitude graphs (Appendix B, Figures B2, B3, B4) are significant in that they give an overall picture of the functions varying in both the order and the argument. This could be useful in many computations, e.g., an integral of the type

$$A = \int_a^b J_n(x)\,dx .$$

Insight might be obtained as to the range of x that would add to the value of the integral and so on.

## REFERENCES

1. Goldstein, M., and Thaler, R. M., "Recurrence Techniques for the Calculation of Bessel Functions," *Math. Tables & Aids to Comput.* 13:102–108, 1959.
2. British Association for the Advancement of Science, "Bessel Functions. pt. II. Functions of Positive Integer Order," Vol. X of series *Mathematical Tables*, Cambridge, England: The University Press, 1931–48.
3. Watson, G. N., "A Treatise on the Theory of Bessel Functions," 2nd. ed., Cambridge, England: The University Press, 1944.
4. Mathematical Tables Project, National Bureau of Standards, "Tables of Spherical Bessel Functions," 2 Vols., New York: Columbia University Press, 1947.
5. Jahnke, E., and Emde, F., "Tables of Functions with Formulae and Curves," 4th ed., New York: Dover Publications, 1945.

## ADDITIONAL BIBLIOGRAPHY

Harvard University, Computation Laboratory, "Tables of the Bessel Functions of the First Kind of Orders," *Ann. Comput, Lab. Harvard Univ.* Vol. 3–14, Cambridge, Mass.: Harvard University Press, 1947–49.

Nosova, L. N., "Tables of Thomson Functions and Their First Derivatives." Translated from Russian by Prasenjit Basu, New York: Pergamon Press, 1961.

Stegun, I. A., and Abramowitz, M., "Generation of Bessel Functions on High Speed Computers," *Math. Tables & Aids to Comput.* 11:255–257, 1957.

# Appendix A

## SOLUTIONS OF BESSEL'S DIFFERENTIAL EQUATION

If $\alpha$ is any real constant, Bessel's equation of order $\alpha$ is

$$y'' + \frac{1}{x} y' + \left(1 - \frac{\alpha^2}{x^2}\right) y = 0 .$$

This is a linear differential equation of second order, and therefore has two and only two independent solutions. If these two solutions are $y_1(x)$ and $y_2(x)$, then there cannot exist two non-zero constants $c_1$ and $c_2$ such that

$$c_1 y_1(x) + c_2 y_2(x) = 0 .$$

That is to say that any other solution, say $y_3(x)$, is a linear combination of the two independent solutions, or

$$y_3(x) = c_1 y_1(x) + c_2 y_2(x) .$$

Note that Bessel's equation has a singular point at $x=0$, but a solution can be obtained by the method of Frobenius. That is, assume a solution of the type

$$y = \sum_{k=0}^{\infty} C_k x^{k+\alpha}$$

and it turns out that the two solution types are completely dependent upon $\alpha$. They are referred to by order $\alpha$ and argument x. If $\alpha$ is an integer, the two independent solutions are $J_\alpha(x)$ and $Y_\alpha(x)$, $\alpha$ = integer. If $\alpha$ is not an integer the two solutions are $J_\alpha(x)$ and $J_{-\alpha}(x)$, $\alpha \neq$ integer. If $\alpha$ is half an odd integer they are usually referred to as the *plus and minus half orders*. $J_\alpha(x)$ is the regular solution; and $J_{-\alpha}(x)$, $\alpha$ not an integer, and $Y_\alpha(x)$, $\alpha$ = integer, are the irregular solutions. The two solutions are also referred to as Bessel functions of the first and second kind respectively.

If the argument x is purely imaginary, the solution is referred to as

$$I_\alpha(x) \quad \text{and} \quad K_\alpha(x) .$$

$I_\alpha(x)$ is analogous to the regular Bessel function and $K_\alpha(x)$ to the irregular Bessel function.

$H_\alpha^{(1)}(x)$ and $H_\alpha^{(2)}(x)$ are known as Bessel functions of the third kind, or first and second Hankel Functions respectively. They are defined as

$$H_\alpha^{(1)}(x) = J_\alpha(x) + i Y_\alpha(x) . \qquad (A1)$$

$$H_\alpha^{(2)}(x) = J_\alpha(x) - i Y_\alpha(x) . \qquad (A2)$$

These are obtained from

$$F_\alpha(x) = A J_\alpha(x) + B Y_\alpha(x)$$

with $A=1$ and $B=\pm i$. If we add and subtract Equations A1 and A2, we arrive at other identities which are sometimes useful:

$$J_\alpha(x) = \frac{1}{2}\left(H_\alpha^{(1)}(x) + H_\alpha^{(2)}(x)\right) .$$

$$Y_\alpha(x) = \frac{-i}{2}\left(H_\alpha^{(1)}(x) - H_\alpha^{(2)}(x)\right) .$$

For complex argument $i^{3/2}x$, one obtains the Thomson function $ber_\alpha(x)$, $bei_\alpha(x)$:

$$J_\alpha\left(i^{3/2}x\right) = ber_\alpha(x) + i\, bei_\alpha(x) .$$

If $\alpha = 0$ the subscripts are usually omitted. Listed below are the two general types of differential equations and their associated functions as

solutions:

$$y'' + \frac{1}{x} y' + k^2 y = 0 \qquad \left\{ \begin{array}{l} y = AJ_0(kx) + BY_0(kx) \\ y = AH_0^{(1)}(kx) + BH_0^{(2)}(kx) \end{array} \right\};$$

$$y'' + \frac{1}{x} y' + \left(k^2 - \frac{\alpha^2}{x^2}\right) y = 0 \qquad \left\{ \begin{array}{l} y = AJ_\alpha(kx) + BY_\alpha(kx) \\ y = AJ_\alpha(kx) + BJ_{-\alpha}(kx) \\ y = AH_\alpha^{(1)}(kx) + BH_\alpha^{(2)}(kx) \end{array} \right\};$$

$$y'' + \frac{1}{x} y' - k^2 y = 0 \qquad \left\{ y = AI_0(kx) + BK_0(kx) \right\};$$

$$y'' + \frac{1}{x} y' - \left(k^2 - \frac{\alpha^2}{x^2}\right) y = 0 \qquad \left\{ \begin{array}{l} y = AI_\alpha(kx) + BK_\alpha(kx) \\ y = AI_\alpha(kx) + BI_{-\alpha}(kx) \end{array} \right\}.$$

# Appendix B

## THE MAGNITUDE GRAPHS

The Magnitude graphs are plots of constant $J_n(x)$ in the x, n plane. The plot of $J_n(x)$ is continuous in this plane both as a function of n and as a function of x. It is a rolling surface and slopes down into the surface of the graph as x and n become larger. Between the zeros, lines of constant magnitude can be drawn, of which there would be two of the same value that connect at some larger x and n value; however, it was felt these would be superfluous here and they were left out. Since the surface is continuous, peak values either plus or minus would fall between two zeros and the magnitude of the function would decrease as we go toward the zero. All lines below the $10^{-2}$ line represent the approximate n and x value for the zeros of the Bessel function with the sign of the function indicated between the zeros.

In looking at the magnitude graphs that follow, pick a particular case, say $J_n(x)$ varying in n for $x=40$; it can be seen that $J_0(40)$ is positive and there are six oscillations before the functions stay positive. It can be seen that

$$-0.3 \leq J_n(x) \leq 0.3 \quad , \qquad 0 \leq n \leq 46 \quad ,$$

$$0 < J_n(x) < 10^{-2} \quad , \qquad n > 46 \quad ,$$

$$0 \leq J_n(x) \leq 10^{-16} \quad , \qquad n > 79 \quad ,$$

and so on. A graph of $J_n(40)$ versus n would appear as in Figure B1.

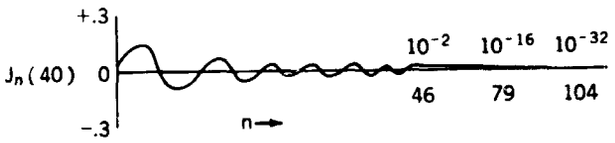


FIGURE B1.—Plot of $J_n(40)$ vs. n.

On the other hand, looking at $J_n(x)$ varying in x, say for $n=50$, the function increases from 0 to $10^{-32}$ for x ranging from 0 to approximately 8.8, and oscillations do not begin until x is approximately 55. It can be seen that

$$J_{50}(x) < 10^{-32} \quad , \qquad 0 \leq x < 8.8 \quad ,$$

$$10^{-32} \leq J_{50}(x) < 10^{-2} \quad , \qquad 8.8 \leq x \leq 44 \quad ,$$

$$-0.3 \leq J_{50}(x) \leq 0.3 \quad , \qquad x > 44 \quad .$$

Therefore, an idea of how the function behaves can be obtained for any order or argument within the limits of the graph.

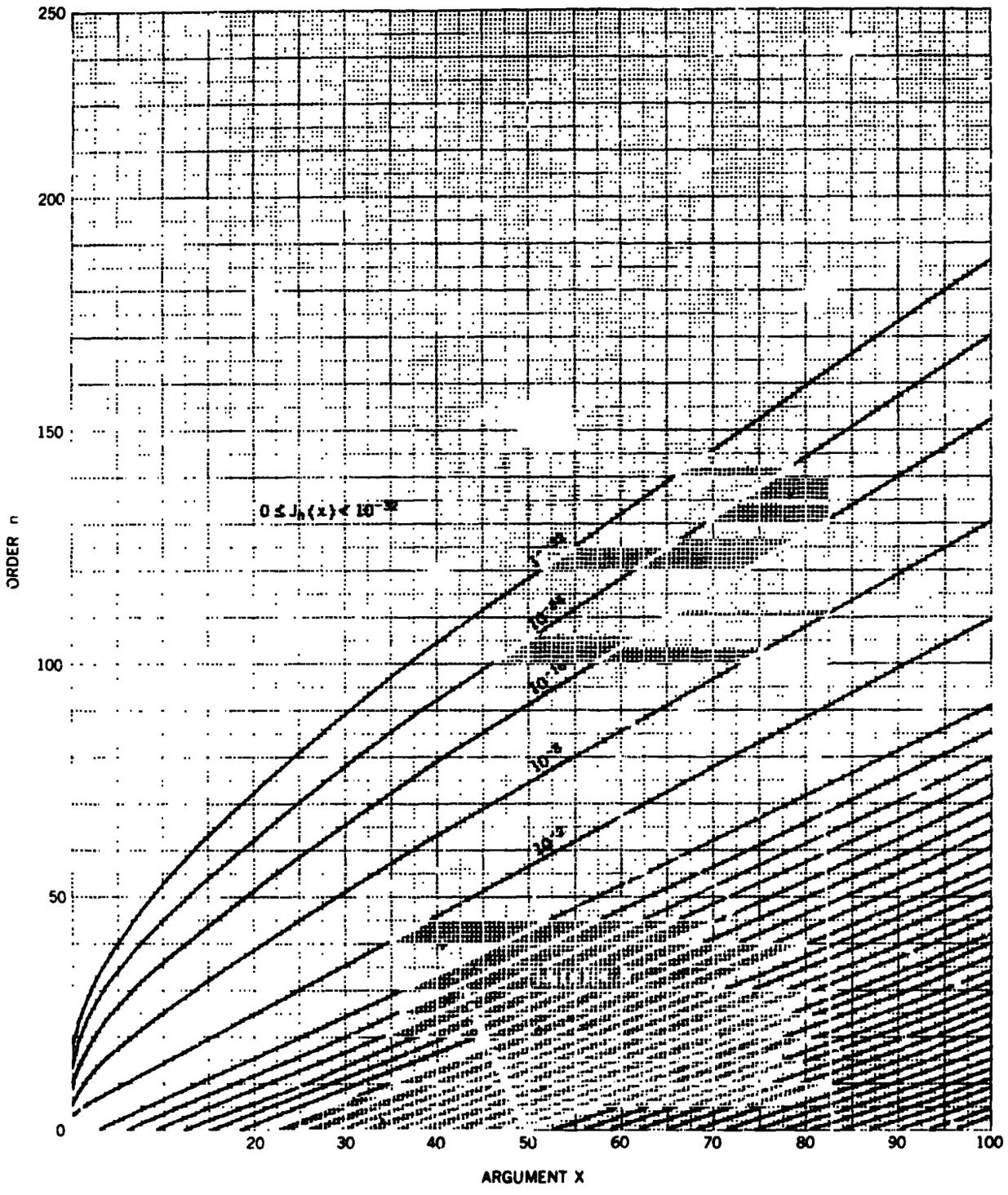


FIGURE B2

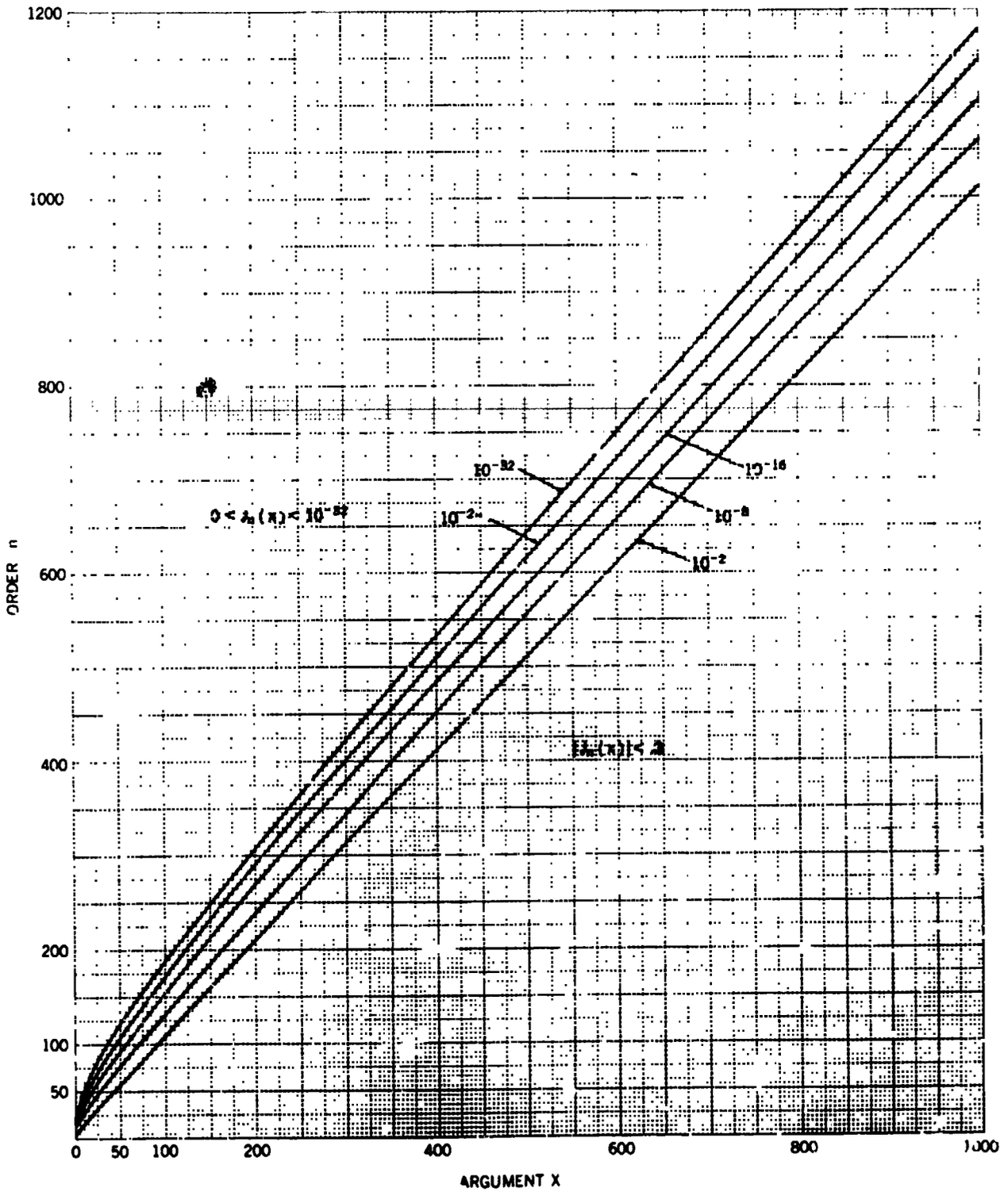
1200
1000
800
600
400
200
100
50
ORDER n
0
50
100
200
400
600
800
1000
ARGUMENT X
$10^{-32}$
$10^{-16}$
$10^{-8}$
$10^{-2}$
$0 < \lambda_n(x) < 10^{-32}$

FIGURE B3

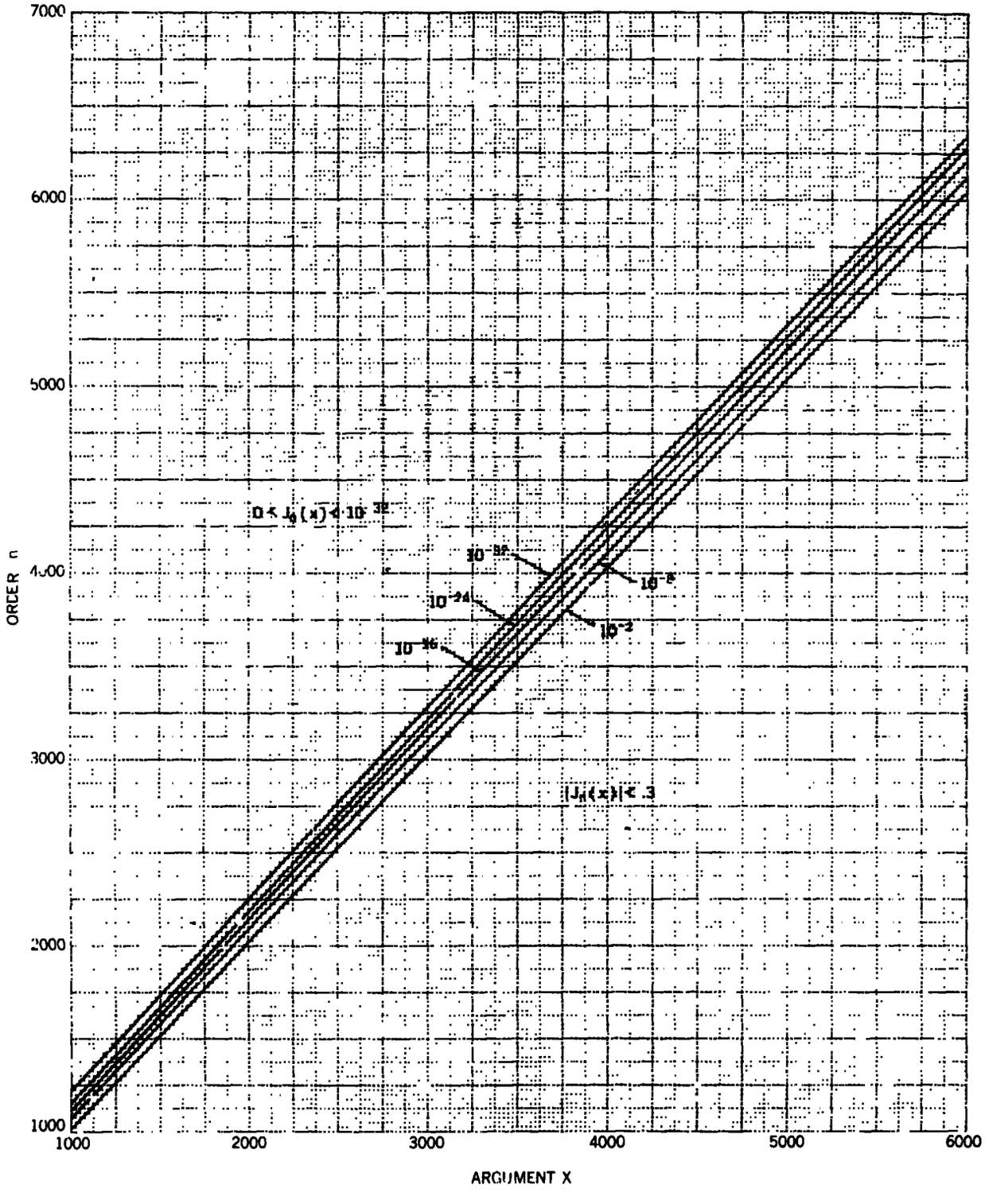
7000
6000
5000
4000
3000
2000
1000
ORDER n
1000
2000
3000
4000
5000
6000
ARGUMENT X
$0 < J_0(x) < 10^{-32}$
$10^{-32}$
$10^{-8}$
$10^{-24}$
$10^{-2}$
$10^{-16}$
$|J_n(x)| < .3$

FIGURE B4