1-1-1977

Light scattering from Saturn's rings calculated by a Markov chain formalism.

Larry Wayne Esposito
University of Massachusetts Amherst

LIGHT SCATTERING FROM SATURN'S RINGS CALCULATED BY A MARKOV CHAIN FORMALISM

A Dissertation Presented

By

LARRY WAYNE ESPOSITO

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

November 1977

Astronomy

LIGHT SCATTERING FROM SATURN'S RINGS CALCULATED BY

A MARKOV CHAIN FORMALISM

A Dissertation Presented

By

LARRY WAYNE ESPOSITO

Approved as to style and content by:

William M. Irvine
Chairman of Committee

Lewis L. House, Member

David J. Van Blerkom, Member

Kari Lumme, Member

George E. McGill, Member

Frederick W. Byron, Jr., Head
Department of Physics
and Astronomy

To My Parents

ACKNOWLEDGEMENTS

It is a pleasure to thank the many people whose assistance and guidance have made this thesis possible. Chief among these are my advisor, William Irvine, and also Kari Lumme and Lewis House, who each supervised a part of this research. I thank David Van Blerkom and George McGill for carefully reading the thesis draft and for their useful comments. I have had helpful discussions with James Cordes, Fred Franklin and Alan Harris. Thanks go to those who assisted with the observations: Holly Ferguson, Don Thompson, Leonard Martin, Helen Horstman, Stuart Jones, Robert Millis, and William Benton. I am grateful to William Baum for his hospitality at the Lowell Observatory and for securing the use of the International Planetary Patrol network for the Saturn's rings observations. Some of the computations were carried out at the National Center for Atmospheric Research, which is supported by the National Science Foundation. I thank Nellie Bristol for her rapid and competent typing of the manuscript. This research was supported in part by NASA NGL 22-010-023.

My special thanks go to my wife, Diane McKnight, who provided encouragement and enthusiasm abundantly throughout the course of this work.

ABSTRACT

Light Scattering from Saturn's Rings Calculated

by a Markov Chain Formalism

(February 1978)

Larry W. Esposito, S.B., M.I.T.

Ph.D., University of Massachusetts

Directed by: Professor William M. Irvine

The theory of Markov chains is used to formulate the radiative transfer problem in a general way by modeling the successive interactions of a photon as a stochastic process. Under the minimal requirement that the stochastic process is a Markov chain, the determination of the diffuse reflection or transmission from a scattering atmosphere is equivalent to the solution of a system of linear equations. This treatment is mathematically equivalent to, and thus has many of the advantages of Monte Carlo methods, but may be considerably more rapid than Monte Carlo algorithms for numerical calculations in particular applications. The speed and accuracy of this formalism have been verified for the standard problem of finding the intensity of scattered light from a homogeneous plane-parallel atmosphere with an arbitrary phase function for scattering. Accurate results over a wide range of parameters were obtained with computation times comparable to those of a standard "doubling" routine. The generality of this formalism may thus allow fast, direct solutions to problems previously soluble only by Monte Carlo methods.

The classical method for accounting for the mutual shadowing among closely packed particles in multiple scattering calculations is extended

in the following ways. 1) By modeling the particle distribution by a Poisson process with a varying density parameter, a "Van der Waals" type approximation allows extension to a greater fractional volume density, D. In this case it is only required that $D^2 << 1$ instead of $D << 1$. 2) In the case that the particle distribution is not uniform the classical calculation may be weighted by the pair correlation function of the distribution. 3) The use of the Markon chain formalism for radiative transfer allows inclusion of the effect of shadowing for two orders of scattering. For conditions such as might apply in Saturn's rings, the inclusion of this effect makes less than 0.1% difference in the calculated phase curves, compared to previous calculations which have included shadowing only in the first scattering. The latter are thus shown to be quite accurate.

Observations are presented of phase curves for both rings in four colors from Saturn's 1977 opposition and linear and logarithmic fits are made to the data. The magnitude of the opposition effect is determined for ring A from photometric data for that ring alone. This appears to have no color dependence at this aspect of Saturn. The opposition effect for ring B decreases with longer wavelength as found by Irvine and Lane (1973). The green phase curve is found to be consistent with earlier observations and is interpreted in terms of multiple scattering models including shadowing, using Markov chain formalism. At least in green, these calculations show that the ring model parameters are consistent with earlier studies of the phase curves and the tilt effect. The differences between the two rings are consistent with the particles having the same scattering properties, so that the two rings need only differ

in optical depth and slightly in volume density. Calculations show that the color dependence of the phase curves can be explained by letting the albedo be a function of wavelength. These calculations support the classical model of Saturn's rings as a layer many particles thick by showing its capability for a consistent explanation of the phase curves over a range of declination of the earth and sun and in all colors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Study of the rings of Saturn has concentrated on the photometric analysis of reflected sunlight. In order to determine the physical properties of the ring particles from these optical observations, two major phenomena must be accounted for. The first is that the solar radiation is multiply scattered, so that an incident photon may be redirected many times before escaping the ring system, or being absorbed within it. The second is that in the classical model of the rings that is many particles thick (e.g., Pollack, 1975) the particles may lie close enough together to cast shadows on each other. In fact, the latter phenomenon is a well-studied explanation of the brightness variation of the rings with solar phase angle (e.g., Kawata and Irvine, 1974). In regard to the first phenomenon, a new method for accounting for the multiple scattering in planetary atmospheres will be developed herein. This will be shown to be of comparable speed and accuracy with standard methods for solving this problem as it applies to Saturn's rings. Several extensions to the classical method of determining the effect of mutual shadowing will be presented and it will be shown that previous analyses are quite accurate. New observations of Saturn's rings from the 1977 opposition will be presented and interpreted in terms of simple multiple scattering models using the formalism of Markov chains. The conclusions are quite consistent with earlier analyses when Saturn was observed at different aspects. A review of the literature pertinent to this study

is now given. For a more general review of Saturn's rings see Pollack (1975) or Palluconi and Pettengill (1974).

The problem of radiative transfer in planetary atmospheres has been treated by numerous authors. The standard text, which is not limited to planetary problems, is by Chandrasekhar (1950). A good review of the methods applicable to planetary problems is that of Irvine (1975). A recent text which is limited to some methods applicable to planetary atmospheres is by Sobolev (1975). For a large number of applications, we can define a "standard problem"; this is to determine the intensity of diffusely reflected light from a plane-parallel homogeneous atmosphere, illuminated monochromatically and monodirectionally from above, and with a perfectly absorbing lower barrier. Ignoring the finite angular size of the sun, this is a good approximation to the rings of Saturn and many of the planetary atmospheres in the solar system. In general, solutions to this problem are compared with the observations to constrain the model parameters. It is much more difficult to solve the "inverse" problem, that is, from the observed intensity to calculate the atmospheric characteristics.

The calculations in this thesis will be compared with those given by the "doubling" routine, which is described by Hansen and Travis (1974). This is based on the principle that if the scattering properties of a single layer are known, we can immediately determine the properties of two such layers placed one atop the other. Starting from very thin layers, we successively "double" the atmosphere until the properties of a layer of the desired thickness are found. This is not the only commonly used method, however, and Lenoble (1977) com-

pares numerous methods and approximations for the solution of the standard problem. Some of the most general routines are based on following the probabilistic development of a single photon, which is sampled in the Monte Carlo methods (House and Avery, 1969). The Markov chain formalism will be based on a similar probabilistic approach. Preisendorfer (1965) and Bharucha-Reid (1960) have shown that various radiative transfer schemes can be formally described as Markov chains. It will be shown later how the basic idea behind the Monte Carlo method, that a photon successively suffers probabilistic interactions, can yield algebraic solutions for the diffusely reflected radiation field based on the formalism of Markov chains. Necessary results from the theory of Markov chains are summarized in the Appendix.

The literature on the effect of mutual shadowing and that on the phase curves of Saturn's rings are intertwined. Improving observations have spurred theoretical advances and vice versa. A good review on these topics is given by Bobrov (1970). The observations of Müller (1893) were the first to show the anomalous brightening of the rings toward opposition which has since been referred to as the "opposition effect". Seeliger (1887) explained this as the result of mutual shadowing among the particles. This work is the basis of the classical method for shadowing calculations which is further discussed in Chapters III and IV. More accurate observations by Guthnik and Prager (1918), Hertzsprung (1919) and Schoenberg (1922) could not be matched by these calculations, however.

As a result, Schoenberg (1933) suggested another explanation. This was that the phase variation was due to single scattering by small par-

ticles. This model concurred with the reported observations of ripples in the phase curves, which were then explained by diffraction. This led to an estimate of 1.8μ as the average radius of the particles, which is substantially smaller than the size required for shadowing to be important.

The work of Seeliger was improved in a series of papers by Bobrov (1940, 1956, 1961, 1970) who removed some of the limitations of the classical theory. He included in his calculations the effect of the finite size of the sun, the contribution of multiple scattering, and the size dispersion of the ring particles. Bobrov also showed that the surface brightness of the rings was too high to be explained by diffraction from spheres, as Shoenberg had suggested.

As found in this work for ring B, Franklin and Cook (1965) and Irvine and Lane (1973) reported that the magnitude of the opposition effect was dependent on color. Including multiple scattering in an approximate way, Franklin and Cook (1965) could explain this by the wavelength dependence of either the glory in the backscattering of an individual particle or of the diffraction of light into the shadows. Both these explanations required small particles, and the latter led to an estimate of the total geometrical thickness of the rings of only 10 cm. This was in conflict with the measurements of Focas and Dollfus (1969) for the thickness of the rings in the range of $\sim$ 2km. However, a reanalysis of those data by Lumme and Irvine (1977) has yielded no lower limit on the actual thickness so that this explanation is no longer ruled out on those grounds. On the other hand, Kawata and Irvine (1974) explained the wavelength dependence as merely due to different

albedos of the particles in the blue and visual (green) portions of the spectrum which led to different contributions of multiple scattering in the two colors. This required a particle albedo in excess of 0.82 at visual wavelengths.

This high value for the albedo was consistent with study of the variation of brightness with changing declination of the earth and sun above the ring plane (the "tilt effect"). Lumme (1970), Price (1973, 1974), and Esposito and Lumme (1977) all concluded from this effect that the albedo in the visual must be at least greater than 0.8 and more likely 0.9. This is required to explain the rapid increase in brightness for ring B with increasing declination as due to increased multiple scattering. Kawata and Irvine (1975) showed that such high albedos could be consistent with the high infrared temperatures observed for the rings (e.g., Murphy, 1974). Lastly, this study is indebted to a recent analysis of photographic data by Lumme and Irvine (1976) for many worthwhile ideas and methods.

Some alternate methods to take mutual shadowing into account have been suggested by Hapke (1963) and Lumme (1971) but they have not achieved wide appeal. This paper will make some extensions to the classical theory of shadowing by including inhomogeneities in the particle distribution, including shadowing in the higher orders of scattering, and allowing applicability to a larger fractional volume occupied by the scattering particles. This last extension requires taking into account the volume occupied by the scatterers themselves to determine their distribution. Although both Seeliger (1887) and Schoenberg (1929) note that this correction must be made, neither states how the correction factors may be

calculated, except that they are very small if the density of particles is low. This correction will be calculated below in a "Van der Waals" type approximation.

In summary, this study will develop a new method for calculating radiative transfer in a planetary atmosphere, apply it to generalizing classical calculations of the effect of mutual shadowing, and use it to interpret new data on the visual phase curves of Saturn's rings.

CHAPTER II

A MARKOV CHAIN FORMALISM FOR RADIATIVE TRANSFER

Introduction

Analysis of observations of a planet in the extended visible region of the spectrum (roughly 0.3 - 1.0 μ) may be complicated by the effect of multiple scattering in the planet's atmosphere. Likewise, the properties of the atmosphere itself may be of interest. In either case, to interpret the properties of a planet or its atmosphere we must be able to account for the processes of interaction between the radiation and the components of the atmosphere. Because the planet is so much colder than the solar photosphere, we can ignore the thermal emission of its atmosphere and merely consider as lost those photons which are absorbed in the atmosphere. The re-emitted photons cannot be confused with those that are merely repeatedly redistributed in direction by conservative interactions with the elements of the atmosphere. This defines a "scattering" atmosphere and poses as the major problem to sum the contributions of light scattered more than once. This may be quite difficult if the redistribution of light at each scattering (the angular redistribution is given by the "phase function") is not isotropic, which is generally the case. The same problem and techniques for solution apply to Saturn's rings. The transfer of the original radiation via multiple scattering gives rise to the observable specific intensity (diffuse reflection) which is measured by instruments on the earth or in space.

A number of methods exist for solving the problem of radiative transfer in a planetary atmosphere (see, e.g., Irvine, 1975). These methods are of varying accuracy, speed, and generality. The most general routines are based on the Monte Carlo method, which follows individual photons as they interact in, and eventually escape from the atmosphere (see House and Avery, 1969). The propagation of each photon is modeled as a stochastic process, and the overall behavior of the radiation field is determined by following the evolution of this process for a large number of photons. The distribution of these trial photons approaches that of light scattered by the atmosphere. Unfortunately, this method can be quite slow because of the need to process many photons to obtain statistically accurate results. However, if the stochastic process satisfies certain general requirements, a solution may be calculated analytically. This results if we model the radiation transport as a finite Markov chain. A Markov chain is a stochastic process having the property that its subsequent evolution depends only on the present state of the system.

With the Markov chain method one can handle problems with more generality than most standard numerical routines. Furthermore, the Markov process is a natural representation for treating a stochastic process such as photon transport.

An example of a complex problem that may be treated by a Markov method is the diffuse reflection of light from a rough surface or atmosphere where the individual scattering elements cast shadows on each other. This generalization has never been included in a natural way in a radiative transfer calculation of multiple scattering, despite the fact that such a treatment is of physical importance in the observation

of non-atmospheric bodies in the solar system such as, for example, the asteroids and the rings of Saturn (Kawata and Irvine, 1975).

In this chapter we take the first step in such a study by determining the capability of Markov chain formalism to provide rapid and accurate solutions for the "standard problem" in radiative transfer in planetary atmospheres, that of determining the diffuse reflection from a plane-parallel, homogeneous atmosphere illuminated by monochromatic, monodirectional solar flux (Chandrasekhar, 1950). We treat the radiation transfer as a Markov chain and compare our solutions for diffuse reflection with those of a standard doubling routine (Hansen and Travis, 1974). At present, the doubling routine is the most common numerical method for solution of light scattering problems in planetary atmospheres.

Markov Chain Formalism

Consider a physical system that evolves in accord with probabilistic laws, that is, a stochastic process. The passage of a photon through a scattering atmosphere can certainly be modeled in this manner. If we include the requirement (the "Markov property") that the relevant probabilities at a given encounter are influenced only by the present state of the photon and not by details of its past history, the stochastic process is a Markov chain (see the appendix for details on Markov chains). By using such a Markov chain as an analogue to the physical process of a photon scattering in an atmosphere, we can mathematically determine the final distribution of radiation leaving the atmosphere. In this section, we will formulate the Markov chain which will serve as

the mathematical analogue of the radiation transport in a scattering atmosphere.

In the present problem, we seek the solution to the following standard problem for planetary atmospheres. We wish to determine the diffusely reflected radiation intensity from a plane-parallel, homogeneous atmosphere illuminated by a collimated, monochromatic solar flux, πF. Our notation follows Irvine (1975): I, specific intensity; τ, optical depth; τ_0, total optical depth; μ, cosine of zenith angle; and ϕ, azimuthal angle. The scattering properties of the medium are specified by an albedo for single scattering, $\tilde{\omega}_0$, and a phase function, P, normalized so that its integral over all directions is 4π. Polarization of the radiation is ignored; there are no internal sources of radiation and the lower boundary is taken to be perfectly absorbing.

The azimuthal dependence is handled by an expansion of all quantities of interest in a Fourier series; see for example, Hansen and Travis (1974). The equation of transfer then reduces to the denumerably infinite set of equations

$$\mu \frac{dI^{(m)}}{d\tau}(\tau,\mu) = I^{(m)}(\tau,\mu) - \frac{\tilde{\omega}_0}{2} \int_{-1}^{1} d\mu' \, I^{(m)}(\tau,\mu') \, P^{(m)}(\mu,\mu')$$

$$m = 0,1,2... \qquad (1)$$

Here, $P^{(m)}$ is the m-th Fourier component of the phase function. The discretization is completed by truncating the series at some finite number, m = M. Many physical phase functions can be expanded in a finite series, in which case this truncation involves no approximation whatever.

We replace the formulation of the continuous range of atmospheric parameters by a discrete analog. The atmosphere is divided into N slabs, termed "layers" or "sub-levels". The n-th slab has upper boundary, $\tau = \tau_n$, and lower boundary $\tau = \tau_{n+1}$. We have $\tau_1 = 0$ and $\tau_{N+1} = \tau_0$. The allowed values of zenith angle are limited to the discrete set of $2N_\theta$ Gauss-Legendre ordinates on the interval $-1 \leq \mu \leq 1$. Thus, the possible values of $\mu = \cos\theta$ form the finite set $\{\mu_1, -\mu_1, \mu_2, -\mu_2, \ldots, \mu_{N\theta}, -\mu_{N\theta}\}$ with $\mu_i \geq 0$. For a given pencil of radiation with zenith angle positive, the radiation is said to be upwelling. Otherwise, the radiation is downwelling.

Thus, the standard problem has been reduced to M azimthally independent radiative transfer problems. The total radiation field is given by

$$I(\tau,\mu,\phi-\phi_0) = I^{(0)} + 2 \sum_{m=1}^{M} I^{(m)}(\tau,\mu) \cos m(\phi-\phi_0) \qquad (2)$$

We solve each of these M problems using a Markov chain formalism. Within the scattering atmosphere a single photon successively suffers interactions, which may result in scattering, absorption or escape from the atmosphere. The Markov chain analogous to this system is defined when we specify its states, the transition probabilities between these states, and the probability distribution for finding the process in any one of the states at some specific time, called the initial distribution. Because the scattering process has the Markov property, once we specify the initial distribution, the future evolution after that time is totally determined. For some details on properties of Markov chains,

see the Appendix.

In the formalism presented here, the state of a photon within the atmosphere is specified by its direction of propagation and the optical depth at the last point of scattering. The range of possibilities represented by attainable states of the Markov chain are labeled by the ordered pairs (i,n), where $1 \leq i \leq 2N_\theta$ and $1 \leq n \leq N$. To say that the Markov process is in a state (i,n) implies that the photon travels in a cone of solid angle $2\pi d\mu_i$ and was last scattered in the n-th sublevel of the atmosphere.

A photon that has escaped from the atmosphere is said to be in state (e) if its propagation is in direction μ_e, where the angles defined by μ_e form a set of emergent rays which may or may not be the same angles as used to discretize the internal radiation field. Another possible fate of a photon is to be absorbed rather than scattered. It then enters a state of the Markov chain that may be labeled (L). Following Preisendorfer (1965), we call it "radiometric limbo".

The attainable states of the Markov chain are: $2 \cdot N_\theta \cdot N$ internal states (i,n); N_e external states (e); radiometric limbo (L). We define a transition between internal states as a scattering event. An internal state passes to an external state by a scattering and subsequent escape without scattering. For a non-conservative interaction, an internal photon is physically absorbed, experiencing a transition to limbo. The transitions of the Markov chain represent the successive interactions of a photon as it passes through the scattering atmosphere. These interactions, which we shall eventually describe as elements of a transition matrix, must account for the processes of physical aborption,

further scattering, or escape.

Once a photon is absorbed or escapes the atmosphere, it suffers no further interactions. Mathematically, this divides the attainable states into two disjoint sets. The internal states are called transient, while the external states and radiometric limbo are ergodic or "absorbing" states. Physically, a photon cannot remain indefinitely in the "transient" states and must eventually reach and remain in one of the "absorbing" states. A chain having this characteristic is an absorbing Markov chain (see the Appendix).

For a finite chain, the probability of reaching any "absorbing" state can be calculated by solving a set of linear equations. We arrange the transition probabilities into an array P, such that P_{ij} is the probability that a photon in state i goes directly to state j. This array (called the transition matrix) can be partitioned as,

$$P = \left[\begin{array}{c|c} I & 0 \\ \hline R & Q \end{array}\right] \quad , \tag{3}$$

where I is the identity matrix; 0 is the null matrix; R is the matrix of transition probabilities from transient to "absorbing" states; Q is the matrix for transition between two transient states (see the appendix). The probability that a photon originally in state i is eventually absorbed by (and thus, remains in) state j is the element X_{ij} of the matrix X which is the solution to the equation,

$$(I-Q)\ X = R \quad . \tag{4}$$

An element of the matrix X is the summation of the probability for ab-

sorption after 1,2,... n transitions as n goes to infinity. Thus, solving this equation corresponds to calculating "all orders of scattering" for the standard radiative transfer problem.

The calculation of the diffuse reflection is given by the following procedure. Discretize the problem so that the resulting stochastic process is an absorbing Markov chain with a finite set of states, as has been described above. Calculate the transition probabilities for each Fourier component, and solve a matrix equation for the absorption probabilities, which gives the Fourier component of the emergent radiation field. Finally, sum the intensity using equation (2) to get the total radiation field.

Because the higher Fourier components of the radiation field arise almost entirely from the radiation that is scattered only once in the atmosphere (Hansen and Travis, 1974) and because the single-scattered intensity may be calculated analytically, a significant gain in speed and accuracy may be achieved by separating the single-scattered radiation from the higher orders of scattering. Thus, we calculate the first-order emergent intensity exactly and use the Markov chain formalism with a Fourier series expansion only for the multiply-scattered light. The emergent intensity is given by

$$\underset{\sim}{I}_{TOTAL} = \underset{\sim}{I}_s + \underset{\sim}{I}$$

$$= \underset{\sim}{I}_s + \underset{\sim}{I}^{(0)} + 2 \sum_{m=1}^{M} \underset{\sim}{I}^{(m)} \cos m(\phi-\phi_0) \quad , \qquad (5)$$

where $\underset{\sim}{I}_s$ is the single-scattered emergent intensity and the vector $\underset{\sim}{I}^{(m)}$

is given by

$$\underset{\sim}{I}^{(m)} = \Pi_0^{(m)} X^{(m)}, \text{ and } X^{(m)} \text{ is the matrix solution to}$$

$$(I-Q^{(m)})X^{(m)} = R^{(m)} \quad . \tag{6}$$

$Q^{(m)}$ and $R^{(m)}$ are the submatrices of $P^{(m)}$ (the Markov chain transition matrix for the m-th Fourier component of the phase function) and $\Pi_0^{(m)}$ is the initial distribution for the Markov chain. Because of the separation of the single-scattered intensity, $\Pi_0^{(m)}$ is the distribution of photons scattered exactly once in the atmosphere.

Calculation of Transition Probabilities

To proceed beyond the formal solution (5), we must specify the transition probabilities, Q and R, and the initial distribution, Π_0, and calculate $\underset{\sim}{I}_s$, the single-scatter intensity at the emergent angles. The calculation of these transition probabilities connects the formal Markov chain theory with the physical system of radiative transfer in a scattering atmosphere. We note that the probability of finding a photon in a given state is proportional to the flux of diffuse radiation classified in that state. For simplicity, we assume that the source function (Chandrasekhar, 1950) is constant over any sublevel. This approximation can be arbitrarily accurate by making the sublevels as thin as necessary.

The state of an internal photon is defined by its zenith angle cosine, μ_i, and the optical depth of last scatter, $\tau \in (\tau_n, \tau_{n+1})$, so that a state k is defined by $k \equiv (i,n)$ and $\ell \equiv (j,n')$. For scattering

between internal (transient) states, the desired transition matrix element is $Q_{k\ell}$ which is the conditional probability, $Q(j,n'|i,n)$. This is the probability that a photon last scattered in sub-level n in direction μ_i, will next be scattered in level n' into direction μ_j.

We will calculate this probability as the product of an extinction probability and a scattering probability. The extinction probability is the probability that a photon from level n, travelling in direction μ_i will suffer extinction (scattering or absorption) in level n'. This probability is independent of the phase function, the incident radiation and the Fourier component. The scattering probability is the probability that the extinction is a scattering in which the scattered photon propagates in direction μ_j. This will be different for each Fourier component.

The extinction probability may be written as the product of three factors. The first is the probability of escape from the layer in direction μ_i. Under the assumption of constant source function, i.e., that the photon is uniformly distributed in (τ_n, τ_{n+1}), this is given by averaging over the level, n:

$$P(\text{escape}) = \frac{1}{\Delta\tau_n} \int_0^{\Delta\tau_n} e^{-\tau'/\mu_i} \, d\tau' = \frac{\mu_i}{\Delta\tau_n} (1-e^{-\Delta\tau_n/\mu_i})$$

where $\Delta\tau_n = \tau_{n+1}-\tau_n$. The second factor is the probability of reaching layer n' from the boundary of layer n, and the third factor is the probability of extinction in layer n'. Multiplication gives the extinction probability $W_{n,i,n'}$ where we distinguish the following cases (note that $\mu_i > 0$):

1) upwelling photons ($n > n'$)

$$W_{n,i,n'} = \frac{\mu_i}{\Delta\tau_n} \left(1-e^{-\Delta\tau_n/\mu_i}\right) \cdot e^{-(\tau_n-\tau_{n'+1})/\mu_i} \cdot \left(1-e^{-\Delta\tau_{n'}/\mu_i}\right)$$

2) downwelling photons ($n < n'$)

$$W_{n,i,n'} = \frac{\mu_i}{\Delta\tau_n} \left(1-e^{-\Delta\tau_n/\mu_i}\right) \cdot e^{-(\tau_{n'}-\tau_{n+1})/\mu_i} \cdot \left(1-e^{-\Delta\tau_{n'}/\mu_i}\right)$$

3) recaptured photons ($n = n'$)

$$W_{n,i,n} = 1 - P(\text{escape}) = 1 - \frac{\mu_i}{\Delta\tau_n} \left(1-e^{-\Delta\tau_n/\mu_i}\right)$$

The scattering probabilities are given by

$$[\text{scatt}]^{(m)}_{ij} = \tilde{\omega}_0 \, P^{(m)}(\mu_i,\mu_j) \, \frac{c_j}{2} \quad ,$$

where $P^{(m)}$ is m-th Fourier component of the phase function and c_j is the quadrature weight for the Gauss-Legendre ordinate μ_j on the interval $[-1,1]$. Thus, we have for the total transition probability between internal states,

$$Q^{(m)}_{k\ell} = W_{n,i,n'} \cdot [\text{scatt}]^{(m)}_{ij} \quad ,$$

where k and ℓ represent ordered pairs (i,n), (j,n') respectively.

To calculate the initial distribution for the Markov chain consistent with the assumption of uniform source function for the transition probabilities, we multiply the probability distribution for photons scattered once within the atmosphere by a correction factor. This is the ratio of the exact single scattered flux escaping the layer to that flux arising from a uniform source of photons within the sublevel. We then have:

1) upwelling photons

$$\Pi_0^{(m)} = e^{-\tau_n/\mu_0} \frac{\mu_i\mu_0}{\mu_i+\mu_0} \tilde{\omega}_0 P^{(m)}(\mu_0,\mu_i) \frac{c_i}{2} (1-e^{-\Delta\tau_n(\frac{1}{\mu_0}+\frac{1}{\mu_i})}) \times$$

$$[\frac{\mu_i}{\Delta\tau_n} (1-e^{-\Delta\tau_n/\mu_i})]^{-1} ,$$

2) downwelling photons

$$\Pi_0^{(m)} = e^{-\tau_n/\mu_0} \frac{\mu_i\mu_0}{\mu_i-\mu_0} \tilde{\omega}_0 P^{(m)}(\mu_0,\mu_i) \frac{c_i}{2} (e^{-\Delta\tau_n/\mu_i} - e^{-\Delta\tau_n/\mu_o}) \times$$

$$[\frac{\mu_i}{\Delta\tau_n} (1-e^{-\Delta\tau_n/\mu_i})]^{-1}$$

unless $\mu_0 = \mu_i$. Then we have

$$\Pi_0^{(m)} = e^{-\tau_n/\mu_i} \tilde{\omega}_0 P^{(m)}(\mu_i,\mu_i) \frac{c_i}{2} [\Delta\tau_n e^{-\Delta\tau_n/\mu_i}][\frac{\mu_i}{\Delta\tau_n} (1-e^{-\Delta\tau_n/\mu_i})]^{-1} .$$

However, this correction introduces an error in the probability that an incident photon is scattered twice in the same sub-level, which is of order $(\frac{\Delta\tau_n}{\mu_i})^3$. This should not be a major source of error if $\Delta\tau_n << 1$.

Next we find the probabilities for escape from the atmosphere, that is, the entries of the matrix $R^{(m)}$, with the same assumptions, by similar integrations over optical depth. We define the probability, $T^1(\mu_j,\mu_e,\tau)$ that a photon incident in direction μ_j on a homogeneous layer of optical depth τ will be diffusely transmitted into direction μ_e after a single scattering.

$$T^1(\mu_j,\mu_e,\tau) = \int_0^\tau \frac{d\tau'}{\mu_j} e^{-\tau'/\mu_j} \tilde{\omega}_0 P^{(m)}(\mu_j,\mu_e) \frac{c_e}{2} e^{-(\tau-\tau')/\mu_e}$$

$$= \frac{\mu_e}{\mu_j-\mu_e} \tilde{\omega}_0 P^{(m)}(\mu_j,\mu_e) \frac{c_e}{2} [e^{-\tau/\mu_e}] \quad .$$

Similarly, the probability for diffuse reflection is given by

$$R^1(\mu_j,\mu_e,\tau) = \frac{\mu_e}{\mu_j+\mu_e} \tilde{\omega}_0 P^{(m)}(\mu_j,\mu_e) \frac{c_e}{2} [1-e^{-\tau(\frac{1}{\mu_j} + \frac{1}{\mu_e})}] \quad .$$

If we assume a uniform source function over the layer n' and that μ_j is upwelling, the transition probability from the transient state $\ell \equiv (n',j)$ to absorbing state (e) is:

$$R_{\ell e} = \int_0^{\Delta t_{n'}} T^1(\mu_j,\mu_e,\tau_{n'}+\tau') \frac{d\tau'}{\Delta\tau_{n'}} \quad .$$

For downwelling μ_j we then have

$$R_{\ell e} = \int_0^{\Delta\tau_{n'}} R^1(\mu_j,\mu_e,\tau_0-\tau_{n'+1}+\tau') e^{-(\tau_{n'+1}-\tau')/\mu_e} \frac{d\tau'}{\Delta\tau_{n'}} \quad .$$

Since we wish to determine the intensity in the ergodic state (e), each of these last transition probabilities is divided by $2c_e\mu_e$.

Lastly, the exact intensity of single-scattered radiation emergent from the atmosphere is given by Chandrasekhar (1950), p. 145.

$$\underset{\sim}{I}_s = \frac{\tilde{\omega}_0}{4} \frac{\mu_0 F}{\mu_e+\mu_0} P^{(m)}(\mu_0,\mu_e)[1-e^{-\tau_0(\frac{1}{\mu_e} + \frac{1}{\mu_0})}] \quad .$$

The values of the above transition probabilities specify the finite Markov chain corresponding to a particular problem of radiative transport. Although our calculations in this section are for a homogeneous atmosphere, by allowing $P^{(m)}$ and $\tilde{\omega}_0$ to vary with optical depth the

formalism would be equally valid for a vertically inhomogeneous atmosphere.

Comparison to Monte Carlo Techniques

As in the Markov chain formalism, Monte Carlo schemes for radiative transfer model the physical interaction as a stochastic process. In both cases, the desired result is the same: they seek the expectation value of the random variable which gives the number of photons emergent from the atmosphere in a particular state. This expectation value is proportional to the specific intensity in that state.

In the Monte Carlo method, the evolution of the stochastic process which serves as a numerical analogue to the physical problem is sampled by drawing random numbers. By the central limit theorem, the statistical characteristics of the sample approach those of the actual stochastic process as the number of elements in the sample becomes large. Specifically, the mean over the sample approaches the actual expectation value of the stochastic process.

A major limitation of Monte Carlo methods is that the convergence is characterized by normal fluctuations of the order $N^{-1/2}$, where N is the size of the sample. Thus, the scheme may require a large amount of computer time if detailed results are desired.

In the Markov chain method, the desired expectation values can be found algebraicly by solving a (large) system of linear equations. In comparison to the Monte Carlo method, this corresponds to taking an infinite number of samples at once. Thus, this calculation yields immediately the limiting distribution to which the corresponding Monte Carlo

calculation converges. Of course, inverting a large number of linear equations can also be a lengthy procedure, so the economy of this method depends on the particular problem. However, the accuracy of the result is limited only by the round-off error in computation and by the degree of correspondence between the numerical analogue and the actual physical system. In the present case, this correspondence is less than perfect because of the discretization in optical depth and cosine of zenith angle. However, every numerical method requires some finite approximation to the physical situation being modeled.

Integral Equations Approach

It is instructive to show how the above linear system for the standard problem can be found from an approach similar to that used for integral equations. If we write a discrete ordinate approximation to the transfer equation after Fourier analysis in azimuth, following Chandrasekhar (1950), p. 56, we get a transfer equation for the specific intensity in each of the discrete directions, i.e.,

$$\mu_i \frac{d}{d\tau} I_i^{(m)}(\tau) = I_i^{(m)}(\tau) - \frac{\omega_0}{2} \sum_j P^{(m)}(\mu_i,\mu_j)\, I_j^{(m)}(\tau) c_j \qquad i = \pm 1 \ldots \pm N_\theta$$

where $I_i^{(m)}(\tau) = I^{(m)}(\mu_i,\tau)$ and c_j is the quadrature weight. If, instead, we rewrite this equation in terms of the upwelling flux in the i-th stream, $F_i^{(m)}$, we have

$$\mu_i \frac{d}{d\tau} F_i^{(m)}(\tau) = F_i^{(m)}(\tau) - \frac{1}{2} \tilde{\omega}_0 c_i \mu_i \sum_j \frac{F_j^{(m)}}{\mu_j}(\tau)\, P^{(m)}(\mu_i\mu_j) \quad .$$

Formal integration yields the coupled Fredholm equations

$$F_i^{(m)}(\tau) = \frac{1}{2} c_i \tilde{\omega}_0 \int_0^{\tau} \sum_j P_{(\mu_i,\mu_j)} \frac{F_j^{(m)}(\tau')}{\mu_j} e^{-(\tau-\tau')\mu_i} d\tau' + f_i^{(m)}(\tau)$$

$$F_i^{(m)}(\tau) = \frac{1}{c} c_i \tilde{\omega}_0 \int_{\tau}^{\tau_0} \sum_j P(\mu_i,\mu_j) \frac{F_j^{(m)}(\tau')}{\mu_j} e^{-(\tau'-\tau)/\mu_i} d\tau' \quad . \qquad (7)$$

The first equation is for downwelling flux, and the second for upwelling flux, and $f_i^{(m)}$ is a forcing function given by the boundary conditions (e.g., solar illumination).

Using a standard approach, we can expand this integral equation in terms of "trial functions". Writing these equations (7) in operator form, we have

$$F = \mathcal{L} F + f \quad .$$

Here, F and f are vectors of $2N_\theta$ components, and $\mathcal{L}$ represents the integral operator in equation (7). We select a finite orthonormal set of functions $\{W_k(\tau)\}$ and perform a generalized Fourier analysis of the system. This yields, for every ℓ,

$$\langle W_\ell, F\rangle = \sum_k \langle W_\ell, \mathcal{L} W_k\rangle\langle W_k, F\rangle + \langle W_\ell, f\rangle \qquad (8)$$

where $\langle W_\ell, F\rangle$ is the scalar product defined by

$$\langle W_\ell, F\rangle = \int_0^{\tau_0} d\tau' \, W_\ell^*(\tau') \, F(\tau') \quad . \qquad (9)$$

This is called "algebraization" of the equation (Green 1969, p. 88). We solve the linear algebraic system

$$F_\ell = \sum_k \mathcal{L}_{\ell k} F_k + f_\ell$$

and construct the solution

$$F(\tau) = \sum_\ell F_\ell \, W_\ell(\tau) \quad .$$

In a recent paper by Cheyney and Arking (1976), a formally similar equation is reached by a variational analysis of the integral equation for the mean intensity. The slight difference arises from their selection of trial functions that need only be linearly independent. Their formalism yields equation (8) when we require the trial functions to be a single orthonormal basis.

The Markov chain method can be considered formally the same as the integral equation approach if we make the following extensions. First, $W_k(\tau)$ is taken to be a unit flux of upwelling (downwelling) radiation in direction μ_i which was last scattered in layer n, i.e., $k \equiv (i,n)$. Also, we define a new scalar product (W_ℓ,F) to be a decomposition of the flux, F, into the portion which was last scattered in layer n' into direction μ_j, $\ell \equiv (j,n')$.

Although these W_k are not orthonormal in the sense of the standard scalar product (9), in the sense of the above decomposition they are both mutually exclusive and exhaustive. The difference can be traced to the fact that in equation (9) the orthonormal bases are vectors in a space of functions while these W_k are functions on a possibility space. We have

$$(W_\ell,W_k) = [W_{(n,i)},W_{(n',j)}] = \delta_{nn'}\,\delta_{ij} \quad .$$

$(W_\ell,F) = 0$ for all ℓ implies $F = 0$. This is sufficient to conclude that the $\{W_\ell\}$ form a complete orthonormal basis with respect to <u>this</u> scalar product. The algebraization now yields

$$F_\ell = \sum_k F_k Q_k + \Pi_\ell \qquad (10)$$

where $Q_{k\ell} = (W_\ell,\mathcal{L}W_k)$ and $\Pi_\ell = (W_\ell,f)$. The choice of notation is not

coincidental: the actual computation of these scalar products is equivalent to the previous calculation of the transition probabilities for the Markov chain. The correspondence is complete if we define the matrix $R_{\ell e}$ which gives the emergent intensity in terms of the internal flux

$$\underset{\sim}{I}_e = \sum_{\ell} F_\ell R_{\ell e} \quad .$$

Solving equation (10), we have

$$F = \Pi(I-Q)^{-1} \quad \text{or}$$

$$\underset{\sim}{I} = \Pi(I-Q)^{-1} R$$

$$= \Pi X \quad .$$

This is seen to be the basic equation of the Markov chain formalism (6). A few comparisons can be made. While a standard integral equation approach requires a decomposition in terms of functions of optical depth, our formalism decomposes the radiation field in terms of its probabilistic development. In an atmosphere which includes mutual shadowing for example, the immediate history of a photon is important, and since the radiation field is not smooth due to the presence of macroscopic scatterers, the Markov chain method would seem preferred.

Mathematically, it is an advantage to work with conserved quantities like probability, as opposed to intensity or mean intensity. The fact that the desired quantities are conserved is sufficient to prove that the linear system is bounded, irreducibly diagonally dominant, and monotone (for proofs see Varga, 1962 and Young, 1971). These properties imply the existence of a unique inverse, and that the system is numerically stable for inversion by a number of methods. This is

quite important in a practical sense, for the ultimate value of a formalism depends on the stability and efficiency with which it can be used to calculate numerical results.

Calculations and Results

A computer code was written to implement a radiation transfer calculation based on this formalism which yields the reflected intensity from a plane-parallel homogeneous atmosphere. The input parameters are the total optical depth of the atmosphere, τ_0; the albedo for single scattering, $\tilde{\omega}_0$; and the scattering phase function for a volume element of the atmosphere, $P(\theta)$. The phase function may be any properly normalized analytic function. A subroutine performs a discrete Fourier transform on the phase function to decompose it into its Fourier components.

This code subdivides the atmosphere into slabs and calculates the Gauss-Legendre ordinates and checks their accuracy. This yields a discretization of the system having a large number of states (typically, 100-200). The transition probabilities between these states are calculated, a matrix inversion is carried out by triangular decomposition, and subroutines perform the necessary matrix multiplications (see equation 6). The inverse Fourier transform yields the total intensity (equation 2).

The results of these calculations have been compared with those from a doubling routine (Hansen and Travis, 1974) as to accuracy, and the timing of the two routines compared for a range of the input parameters. In all cases, the phase function was taken to be a Henyey-

Greenstein function,

$$P(\theta) = \frac{(1-g^2)}{(1+g^2-2g\cos\theta)^{3/2}} \quad .$$

This allows a large variety of cases to be studied by varying a single parameter, g.

The comparisons are carried out for the range

$$0.01 \leq \tau_0 \leq 4.0$$
$$-0.7 \leq g \leq 0.5$$
$$0.1 \leq \tilde{\omega}_0 \leq 1.0 \quad .$$

A set of computations are summarized in Table 1. Markov chain calculations were performed with the atmosphere divided into five, ten, and twenty sublayers. The highest accuracy for a given number of divisions results from dividing the atmosphere unequally, so that the layers nearer the boundary were progressively thinner. The number of Gauss-Legendre ordinates in the zenith angle was taken to be five, six, ten, or twelve.

In the results that follow, we define the relative error, ε, as the maximum relative difference between the diffusely reflected intensity in the Markov chain calculation and the doubling routine:

$$\varepsilon = [I(\text{doubling}) - I(\text{Markov})]/I(\text{doubling}) \quad .$$

The intensity was compared for $\mu_e = \mu_0$, $\alpha = 6°$; $0.07 \leq \mu_0 \leq 0.9$, where α is the phase angle between observer and the source of illumination. Such geometric configurations are characteristic of earth-based observations of the outer planets.

Excellent agreement is found for the easiest cases of $\tau_0 << 1.0$, $\tilde{\omega}_0 << 1.0$ (eight significant figures). This corresponds to the same

accuracy to which the Gauss-Legendre ordinates and weights are calculated in each routine so that better agreement is hardly possible. These results give us considerable confidence in the correctness of the formalism.

For more interesting cases, we find the relative error to scale as $(\tilde{\omega}_0 \Delta\tau)^2$, where $\Delta\tau$ is a characteristic optical depth in a single sublayer. For a given physical situation, this means that the error decreases as $1/N^2$, where N is the total number of layers in the optical depth subdivision. This behavior is understandable because the formalism cannot follow explicitly the photons that are scattered more than once in a given layer. The proportion of radiation which is multiply scattered in a layer is just of the order $(\tilde{\omega}_0 \Delta\tau)^2$.

For light emerging at small μ (grazing reflection), the relative error is considerably reduced by using an unequal subdivision of the atmosphere. The fact that the layers are thinner near the surface compensates for the more rapid variation of the radiation field near the boundary. The following results all represent calculations with the layer boundaries distributed according to

$$\tau_n = \tau_0\{-\ell n\ [1 - (\frac{n-1}{N\sqrt{3}})]\} \quad .$$

It is important to know how the numerical implementation of this formulation compares with the standard routines for calculating radiative transfer in a scattering atmosphere. Since the accuracy of the numerical inversion is limited only by the round-off errors, the accuracy of this method depends almost entirely on the coarseness of the approximation by a finite mathematical analogy to the actual continuous

system. One would expect that by making the number of states arbitrarily large, the relative error could be made arbitrarily small. Our calculations, summarized in Table 1, support this. We find that a calculation based on a Markov chain with 60 states gives a maximum relative error of less than 1% for isotropic scattering atmospheres with $\tau_0 \leq 1.0$, $\tilde{\omega}_0 \leq 1.0$. For a chain of 120 states, the same accuracy could be achieved over the range $-0.5 \leq g \leq 0.4$ with $\tau_0 \leq 2.0$, $\tilde{\omega}_0 \leq 0.9$ or $\tau_0 \leq 4.0$, $\tilde{\omega}_0 \leq 0.6$. For a chain of 200 states, the 1% error range included $-0.7 \leq g \leq 0.5$, $\tau_0 \leq 4.0$, $\tilde{\omega}_0 \leq 0.9$.

The amount of computer time required for these calculations scales roughly as the cube of the total number of states. This reflects the fact that the majority of this time is spent in the triangular decomposition of the linear system, which requires $N^3/3$ arithmetic operations. A Markov chain calculation with 120 states required very nearly the same amount of time as the doubling ($\pm 5\%$). The calculation with 200 states required four times as much.

The fact that the two routines are comparable in time is in part due to the effectiveness of separating off the single-scattered intensity (equation 5). Because of this, only about half as many Fourier terms need be calculated as are necessary to describe the total radiation field, including the single scattering.

Because the majority of the computing time is spent in solving the system of linear equations, it seems reasonable to consider other strategies besides triangular decomposition to accomplish this. For example, the matrix inversion may be calculated iteratively by a Neuman series expansion (see Green, 1969, p. 31f), which formally corresponds to the

Table 1

Accuracy of the Markov chain formalism. N is the number of divisions in optical depth. Time is the ratio of CPU time to that for a "doubling" calculation.

τ_o	$\tilde{\omega}_0$	g	Total States	N	ε	Time
0.01	1.0	0.0	60	5	$1x10^{-8}$	0.6
0.1	1.0	0.0	60	5	2x10-8	0.6
0.5	1.0	0.0	60	5	$2x10^{-4}$	0.6
1.0	0.9	-0.4	60	5	$1x10^{-3}$	0.6
2.0	0.9	0.5	120	10	$2x10^{-2}$	1.0
2.0	0.9	-0.5	120	10	$5x10^{-3}$	1.0
3.0	1.0	0.0	120	10	$2x10^{-2}$	1.0
4.0	0.9	0.5	200	20	$3x10^{-3}$	1.5
4.0	0.9	-0.7	200	20	$4x10^{-3}$	1.5

radiative transfer method of "successive orders of scattering" (Irvine, 1964). Radiative transfer calculations using an algorithm based on this iterative approach were compared with the above calculations. Not surprisingly, the numerical accuracy was unaffected by this procedure since we required the Neumann series to converge to 5 parts in 10^5. In regard to timing, this method was in fact 40% slower than the direct method for a chain of 120 states. However, because the number of operations required for the Neumann series iteration scales as n^2 (as opposed to n^3 for triangular decomposition), the calculation with 200 states required the same time for both methods. A greater gain in speed was accomplished by combining the two methods. When the lowest Fourier components were calculated directly and the Neumann series was used for the higher components, the calculation with 200 states required only 50% more time than the doubling routine, that is, it was three times faster than using either the triangular decomposition, or Neumann series iteration alone.

Conclusions

The preceding calculations and computations indicate that a radiative transfer calculation based on a Markov chain formalism can be both accurate and economical. For scattering problems with $-0.5 \leq g \leq 0.4$, $\tau_0 \leq 2.0$, $\tilde{\omega}_0 \leq 0.9$, this formalism yields results accurate to better than 1% in the same time required for a doubling calculation. By expanding the number of states in the Markov chain, this error can be made smaller, scaling roughly as $1/N^2$, where N is the number of divisions in optical depth. For a calculation using 200 states, the region

with 1% accuracy included $-0.7 \leq g \leq 0.5$, $\tau_0 \leq 4.0$, $\tilde{\omega}_0 < 0.9$.

Despite this, for the standard case of a plane-parallel, homogeneous atmosphere, the doubling routine is superior because of its speed and applicability over a much wider range of parameters, especially for large optical depths. The advantage of the Markov chain formalism, however, is that it can be easily generalized to include vertical inhomogeneities and complicated geometries. However, since the requirements of computer storage increase with the complexity of the problem, the economy of the numerical implementation of this formalism remains to be seen in these cases. It has many of the advantages of the Monte Carlo method, allowing a finite representation of a scattering system in terms of its probabilistic evolution. This requires only that the propagation of radiation can be modeled as a stochastic process having the Markov property. In the following chapter, the inclusion of the shadowing among macroscopic elements in an atmosphere will be treated in this way, extending existing calculations. This, and other problems previously tractable only by Monte Carlo methods may now be studied with a new method where the limiting distribution of radiation can be calculated algebraicly.

CHAPTER III

EXTENSIONS TO THE CLASSICAL CALCULATION OF THE EFFECT OF MUTUAL SHADOWING IN DIFFUSE REFLECTION

Introduction

In some physical situations the scatterers in a medium may be large enough and close enough to cast shadows on each other. This situation will hold whenever geometric optics may be used to describe the photon propagation between successive scatterings (in this case it is meaningful to speak of shadows) and the particle size is comparable to the interparticle separation. If this is the case, the transfer of radiation is not well described by the equation of radiative transfer. As a result, the distribution of radiation scattered by the medium is not given by the standard solutions to the transfer equation. The difference, most pronounced at small phase angles where the particles may hide their own shadows, is called the "shadowing effect". This effect has been proposed as an explanation for the phase curve of the Moon and also Saturn's rings (e.g. Irvine, 1966).

Classically, the problem of calculating the shadowing effect has been studied by a number of authors (e.g., Seeliger 1887; Schoenberg, 1929; Bobrov, 1940, 1961, 1970; Hapke, 1963; Irvine, 1966; Lumme, 1970). In all of these calculations, it is assumed that the fractional volume occupied by the scatteres, D, is much less than unity. Since the effect increases with D, extensions that are valid for larger values of this density are clearly desirable. It is also assumed that the location of

the particle centers is unaffected by any mutual interactions so that their distribution is statistically uniform. Recently, Franklin and Colombo (1977) and Colombo, Goldreich and Harris (1977) have suggested that the observed brightness variation in the A ring of Saturn may be explained by assuming that the distribution of scattering particles is not uniform due to mutual gravitational perturbations. This complication needs to be taken into account in matching radiative transfer models of these azimuthal variations with dynamical analysis and observations of Saturn's rings. Further, the classical calculations only account for the effect of shadows on the radiation that is scattered just once. The validity of these calculations to study the light diffusely reflected from astronomical bodies depends on the error made in ignoring shadowing in the higher orders of scattering. Irvine (1966) has argued that this error is small, but no explicit calculations have shown the magnitude of this effect for solar system objects. By considering higher orders of scattering in terms of a Markov chain, we can make an estimate of this error. This paper will demonstrate generalizations of the existing methods in which the above limitations can be at least partially removed.

Extension to Greater Density

Seeliger (1887) was the first to formulate the classical method for accounting for mutual shadowing. A good review of this approach is given by Irvine (1966). Because of the assumption of geometric optics, Irvine notes that these calculations are relevant when we have

$$\lambda \ll \rho^2/\Delta \quad , \tag{11}$$

where λ is the wavelength of the photon, ρ is the particle radius, and Δ is the lesser of the mean free path for a photon between scatterings and the total extent of the scattering layer. If in the opposite extreme, we have either

$$\rho^2/\Delta \ll \lambda \ll \rho \quad \text{or} \quad \lambda \geq \rho \quad , \tag{12}$$

it is not meaningful to speak of shadows and the standard methods of radiative transfer apply. (i.e., the scattering centers are in each others far field.)

Classically, the effect of mutual shadows is determined in a single scattering geometry by calculating the total volume, V, of the layer which must be void of scattering centers for a ray to penetrate to a given depth, scatter, and escape. A calculation of this volume will be given in the next subsection. Once this volume may be calculated as a function of the scattering geometry, an integration over the layer gives the primary scattered component of the diffuse reflection. At the surface of the atmosphere this intensity is given by (the asterisk denotes the inclusion of shadows)

$$I_1^* = \frac{\tilde{\omega}_0}{4\mu} P(\Omega_0,\Omega) \int_0^{\tau_0} d\tau' \, \underline{P}(k=0,V) \quad . \tag{13}$$

Here $\tilde{\omega}_0$ is the albedo for single scattering; $\Omega_0 = (\theta_0,\phi_0)$ is the incident angle and $\Omega = (\theta,\phi)$ is the scattering angle; $\mu = \cos\theta$ is the cosine of the zenith angle for the scattered ray; P is the phase function for scattering; and τ_0 is the total optical depth of the layer. An incident solar flux of unity has been assumed. $\underline{P}(k=0;V)$ is the probability that no particle (k=0) lies in the volume V. The dependence of V on volume

density D, phase angle α, τ', μ_0 and μ (which is a later calculation) has been suppressed.

Classically, the desired probability is given by assuming a Poisson distribution for the particle centers, i.e.

$$\underline{P}(k=0;V) = \exp(-n_0 V) \quad , \tag{14}$$

where n_0 is the number density of scatterers in the medium. If the volume of a particle is v_1, then we have

$$D = n_0 v_1 \quad . \tag{15}$$

If D is significantly larger than zero, the assumption of Poisson distribution will be in error, as it gives a finite probability that scatterers overlap, a physical impossibility. This difficulty is similar to that of the ideal gas law at higher density. As in statistical mechanics we can hope to make a first-order improvement to the description of the particles by a "Van der Waals" type approximation.

Formally, we relax the requirement of Poisson distribution while maintaining the requirement of uniformity in the following form: provided the region of interest lies outside of the volume known to be occupied by a particle, the probability of encountering a particle is constant throughout the medium and independent of the locations of particles more distant than the dimension of a single particle. This will be a good approximation to situations where the forces between particles are very weak. In essence, this requirement limits the interactions of the particles to "hard sphere" forces which exclude any two particles from overlapping. Given these assumptions, it is the encounters of a photon along a given path which are distributed with uniform likelihood.

A mathematical model of this is provided by a Poisson process with a varying density parameter. We define $\underset{\sim}{\lambda}(t)$ to be the probability density of an encounter at the point t for a photon traveling some path containing t. The above assumptions of uniformity imply

$$\begin{aligned} &\underset{\sim}{\lambda}(t) = \underset{\sim}{\lambda}' = \text{constant} \qquad \text{outside the particles} \quad , \\ &\underset{\sim}{\lambda}(t) = 0 \qquad\qquad\qquad\quad \text{within a particle.} \end{aligned} \tag{16}$$

The mean number of interactions (scattering or absorption) along a ray path of length h is given by the line integral

$$\langle k \rangle = \int_0^h \underset{\sim}{\lambda}(t)\, dt \quad . \tag{17}$$

Because of the uniformity of the medium, any line segment has a fraction D that is covered by particle cross-sections. This yields

$$\langle k \rangle = (1-D)\, h\underset{\sim}{\lambda}' \quad . \tag{18}$$

But this must also equal the expectation value of the number of particles along the given path, i.e., the mean number of particles in a cylinder of cross-section $\langle A \rangle$ and length, h. Here, $\langle A \rangle$ is the mean extinction cross-section for the scattering particles (for spheres, this is $n\rho^2$ in the geometric optics approximation). This gives

$$\begin{aligned} &(1-D)h\underset{\sim}{\lambda}' = n_0 \langle A \rangle h \quad , \quad \text{or} \\ &\underset{\sim}{\lambda}' = \frac{n_0}{1-D} \langle A \rangle \quad . \end{aligned} \tag{19}$$

For a Poisson process, the probability of k events (encounters, extinctions) in the interval [0,h] is given by (e.g., Papoulis, 1965)

$$\underline{P}(k;h) = \frac{(\int_o^h \underset{\sim}{\lambda}(t)\, dt)^k}{k!} \exp(-\int_o^h \underset{\sim}{\lambda}(t)dt) \quad . \tag{20}$$

The important probability for scattering calculations is for $k = 0$, as in (14). In this case, no particle cross-sections are encountered, so that $\underset{\sim}{\lambda}(t) = \underset{\sim}{\lambda}'$. This gives

$$\underline{P}[k=0;h] = \exp\left(\frac{-n_o}{1-D} <A> h\right) \quad . \tag{21}$$

Compare this with (14) with $V = <A>h$ which yields

$$\underline{P}[k=0;V] = \exp\left(-n_o <A> h\right) \quad . \tag{22}$$

It can be seen that the classical formalism may be extended to greater particle density, D by replacing n_0, the actual number density by

$$n' = \frac{n_o}{1-D} \quad . \tag{23}$$

In this extension, D enters in the first order. This approximation thus ignores terms of second order and higher in D, so that its validity is limited to $D^2 << 1$. Such terms would arise from mutual interactions among the particles, n body interactions having a dependence of D^n.

Physically, the extension is familiar from part of the method of reaching the Van der Waals equation of state from the ideal gas law. N is the total number of particles, each with volume v_1, in a large volume V. The Van der Waals approximation takes into account the finite size of the particles by considering the "effective" volume $V-Nv_1$ available to the particles. This gives an effective number density

$$n' = \frac{N}{V-Nv_1} = \frac{n_o}{1-D} \quad , \tag{24}$$

as above. In the limit of low D, n' goes to n_0, yielding the classical

result in equations (14) and (22).

Extensions to Non-uniform Media

A desirable extension to the classical theory is to include the possibility of non-uniformity in the scattering medium. A simple way to describe the departure from uniformity is in terms of the pair correlation function. This function gives the relative likelihood of finding a particle as a function of the distance from a known particle. For uniform media this function is merely unity. Despite the simplicity of description by a single function, this approach has wide applicability. It yields an accurate description of the statistical mechanics of quite dense liquids (e.g., Cole, 1959). Further, calculations of the effects of mutual gravitational interactions in Saturn's rings have yielded an estimate of the pair correlation function for the ring system particle distribution (Alcock and Goldreich, 1977). It will be shown how this function can be used to correct the classical formulation by including non-uniformity.

In order to incorporate this additional complication, we will replace the volume integrals of earlier formulations by path integrals with a definition of the instantaneous cross-section for extinction (including shadowing) at each point along the path. In addition, it will be necessary to factor this instantaneous cross-section into two parts, a purely geometrical part (independent of particle distribution), and the local density given by the pair correlation function which describes the non-uniformity.

The best starting point for this discussion is the classical dia-

gram showing the geometry of the shadowing effect. See Figure 1. The following discussion will assume identical spheres as the scattering particles. This is both for simplicity and because the maximum magnitude of the shadowing effect at $\alpha = 0°$ can be shown to be independent of particle shape, and to depend only on the average cross-section for extinction (Seeliger, 1895). In addition, Bobrov (1961) has shown how a dispersion in particle size may be taken into account.

If the scattering particles are spheres of radius ρ which is much less than the vertical extent of the scattering layer, then the volume V^- is a cylinder of volume $V^- = \frac{\tau'}{n_0\mu_0}$ and V^+ has volume $V^+ = \frac{\tau'}{n_0\mu}$. The total volume which must be void of particle centers (as discussed above) for a photon to enter and escape along the shaded path is

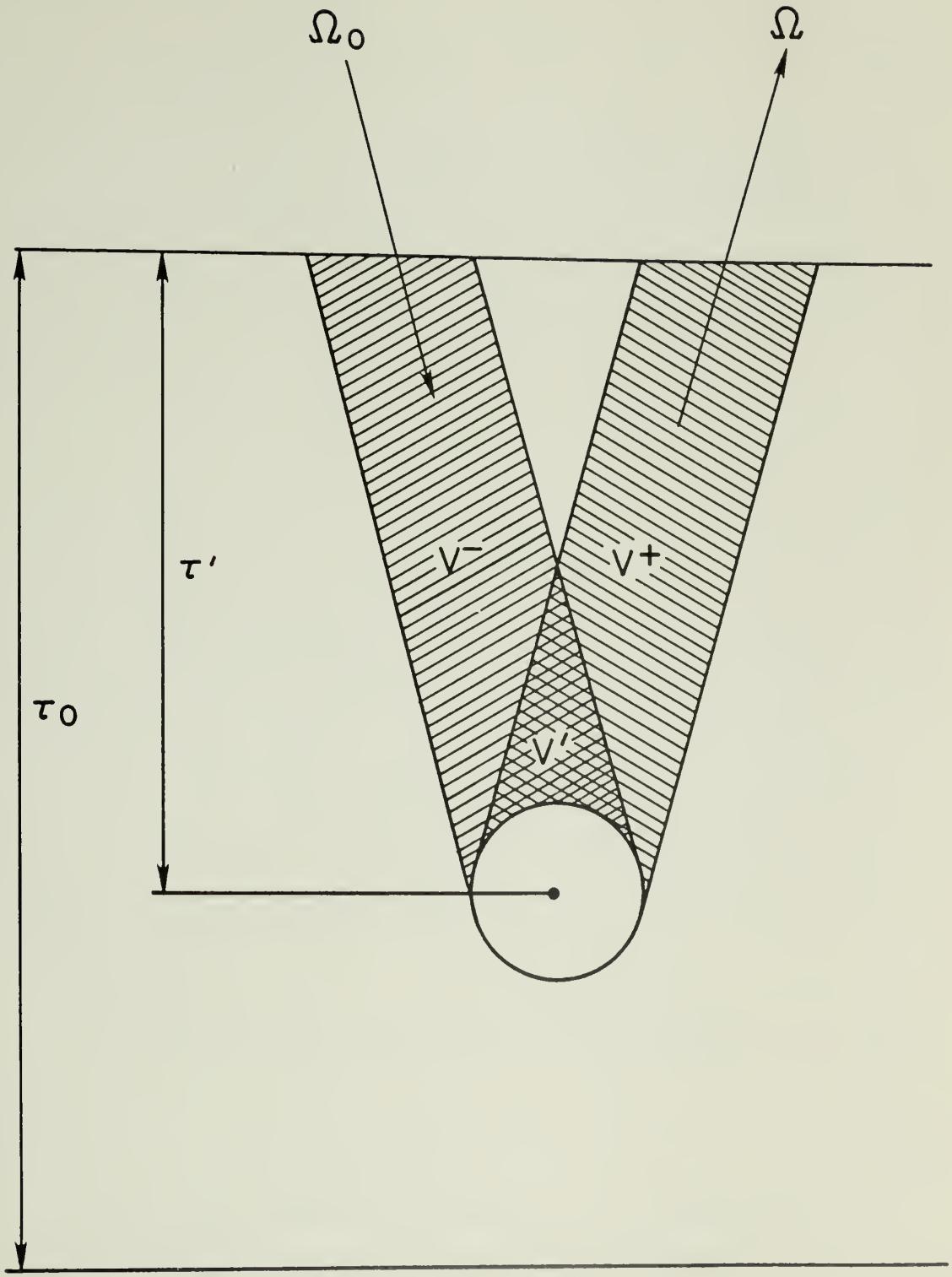
$$V = V^- + V^+ - V' \quad . \tag{25}$$

It is the calculation of this volume of overlap, V', which determines the magnitude of the shadowing effect. When V' goes to zero there is no shadowing effect and the standard methods of radiative transfer will apply. When shadowing plays a role, we can clearly see how this quantity is important by writing equation (13) as

$$\begin{aligned} I_1^* &= \frac{\tilde{\omega}_0}{4\mu} P(\Omega_0,\Omega) \int_0^{\tau_0} d\tau' \exp[-n_0 V] \\ &= \frac{\tilde{\omega}_0}{4\mu} P(\Omega_0,\Omega) \int_0^{\tau_0} d\tau' \exp[-\tau'(\frac{1}{\mu_0}+\frac{1}{\mu}) + n_0 V'] \quad . \end{aligned} \tag{26}$$

The volume V' may be found by evaluating a volume integral, as in Irvine (1966). However, let us replace the quantity n_0V' by a path integral along the outgoing ray (axis of the cylinder V^+ in Figure 1). If the geometric distance traveled by the photon is given as $h = \tau'/(n_0\pi\rho^2\mu)$,

Figure 1. Shadowing geometry after Irvine (1966). Photon enters from Ω_0 and is scattered at τ' into direction Ω.


Ω_0
Ω
τ'
τ_0
V^-
V^+
V'

then we can write

$$n_0 V' = \int_0^h dt\, n(t)\sigma(t) \quad . \tag{27}$$

We define $n(t)$ to be the local density at the particular point t and $\sigma(t)$ is thus the instantaneous correction to the cross-section for extinction which is due to shadowing. The dependence of $\sigma(t)$ on phase angle and volume density has been suppressed. If the medium is uniform, then we have

$$V' = \int_0^h \sigma(t)\, dt. \tag{28}$$

The quantity $\sigma(t)$ is thus the area common to both cylinders V^- and V^+ in a plane perpendicular to the outgoing ray at the point t. Following Bobrov (1961) and Franklin and Cook (1965) we will assume that the ellipse formed by the cross-section of the cylinder V^- in this plane may be approximated by a circle. This will be an excellent approximation in the present case because shadowing is only important when the phase angle, α, is very small. We note that the eccentricity of the ellipse in question is just $\sin\alpha$ which vanishes in the same limit. We find (for equal spheres)

$$\sigma(t) = \pi\rho^2 - 2T(\rho^2 - T^2)^{1/2} + 2\rho^2 \sin^{-1}(Tp) \quad , \tag{29}$$

where

$$T = t \tan(\alpha/2) \quad . \tag{30}$$

We have checked this approximation for the uniform density case where the scattering particle is far enough from the layer boundary so that the entire volume of overlap lies within the atmosphere. This gives the maximum value of V' for a given phase angle:

$$V' = \int_0^{h=\rho \cot(\alpha/2)} dt\ \sigma(t) \quad .$$
$$= \frac{4}{3}\rho^3[\cot(\alpha/2)] \quad . \tag{31}$$

This agrees exactly with the value found without approximation by Irvine (1966), correcting an obvious error in the latter. Further, in a numerical calculation of light scattered by Saturn's rings we have compared our values from the path integral method with those tabulated by Schoenberg (1929). The two calculations agree to one part in 10^4 in the range $0.1° \leq \alpha \leq 6.5°$.

To determine n(t), it is assumed that the departure from uniformity is described by a single function of the distance from the last scattering. The pair correlation function $g(\vec{r})$ give the mean particle density at a position $\vec{r}$ relative to the average probability of finding a particle at any location; i.e.,

$$n(t) = n_0\, g(\vec{r}) \quad . \tag{32}$$

This single function is in any case the simplest characterization of non-uniformity in the medium.

With these assumptions, we can calculate the effect of shadowing by the following replacement in equation (16)

$$n_0 V' = \int_0^h dt\ n_0\, g(\vec{r})\sigma(t) \quad . \tag{33}$$

If the fractional volume occupied by scatters is not negligible, then $g(\vec{r})$ should be replaced by its "Van der Waals" analog:

$$g(\vec{r})/(1 - g(\vec{r})D) \quad , \quad \text{as above.}$$

These extensions now allow calculation of the effect of mutual shadowing in non-uniform media where $D^2 << 1$, but not necessarily $D << 1$. For uniform media with small volume density, this generalization reproduces the classical result.

Extension to Higher Orders of Scattering

Once the effect of shadowing is given for single scattering, the total diffuse reflectivity may be calculated as follows (Irvine, 1966). The solution of the transfer equation is determined in the absence of shadowing. The intensity of radiation is expended in a Neumann series (expansion in successive orders of scattering) so that we have (e.g., Irvine, 1965).

$$I(\tau,\Omega) = \sum_{n=0}^{\infty} (\tilde{\omega}_0)^n I_n(\tau,\Omega) \quad . \tag{34}$$

Here, $I(\tau,\Omega)$ is the specific intensity at depth τ in the direction Ω; $\tilde{\omega}_0$ is the albedo for single scattering; and $I_n(\tau,\Omega)$ is the intensity of photons scattered exactly n times in a conservative atmosphere. The effect of shadowing is taken into account by replacing $\tilde{\omega}_0 I_1(\tau,\Omega)$ by the single scattered intensity including shadowing (as given by equation (26), for example). This gives (asterisks implying the inclusion of the effect of mutual shadowing)

$$I^*(\tau,\Omega) = I_1^*(\tau,\Omega) + \sum_{n=2}^{\infty} (\tilde{\omega}_0)^n I_n(\tau,\Omega) \quad , \tag{35}$$

where I_1^* is given by equation (26) or one of the similar calculations referenced in the introduction.

The shortcomings of this classical approach are: 1) shadowing is

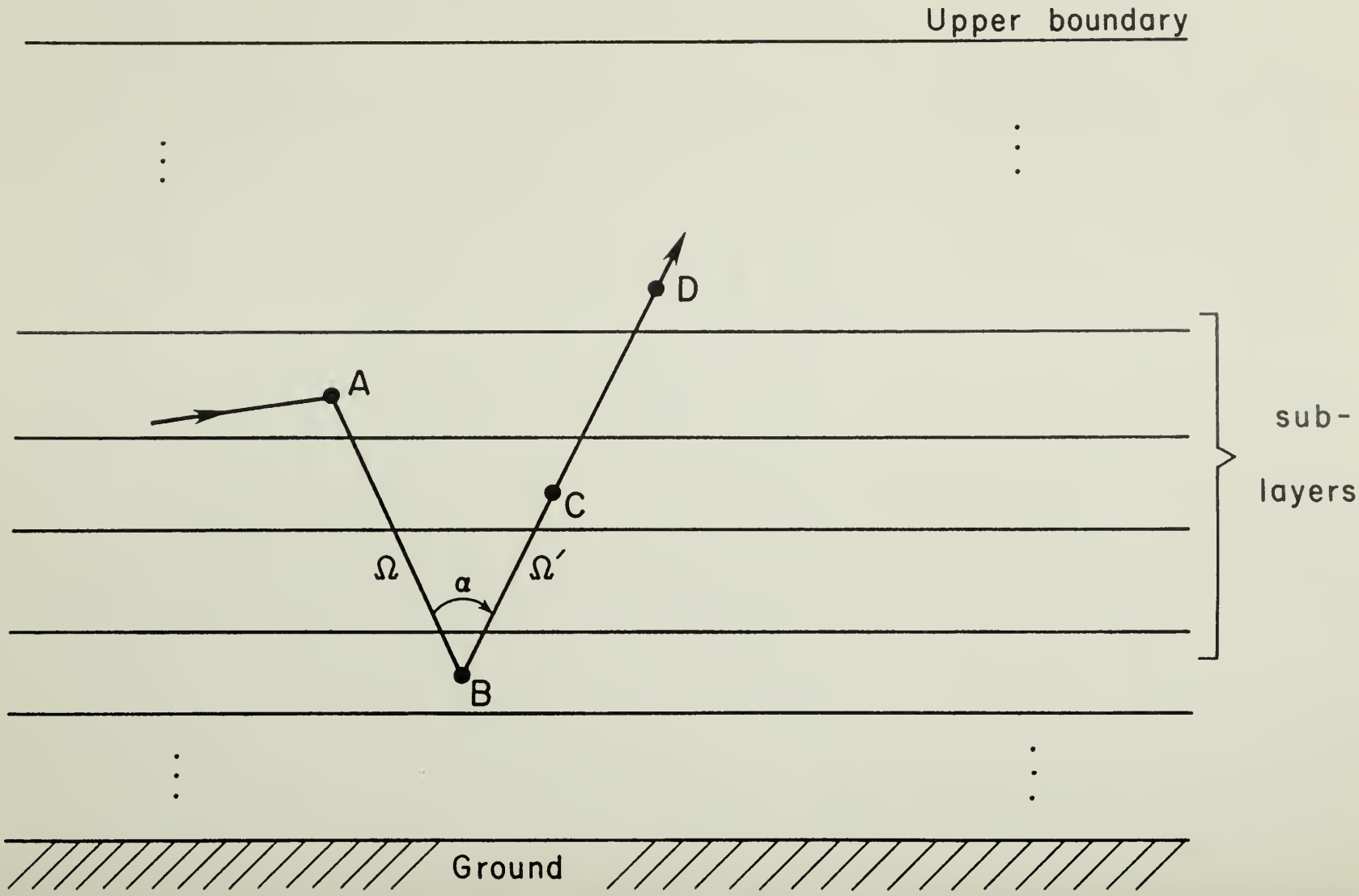
not taken into account in any scattering except the first and 2) if we consider the higher order terms being found iteratively from the lower, these are calculated by successive iteration on I_1 and not the true single scattering I_1^*. Regardless of the effect of shadowing in photon redistribution at each scattering, the higher orders of scattering will be inconsistent with the first. It is possible to improve on the classical method by treating the radiative transfer as a stochastic process. In the formalism based on Markov chains developed in the previous chapter, the distribution of photons diffusely reflected by a scattering layer is determined from only the single scattering properties. The Markov property and the probabilities for single scattering transitions suffice to yield the radiation field including all orders of scattering.

In practice, this procedure has certain difficulties. The major one is as follows: while in the situation without mutual shadowing the probability for scattering at any given point depends only on the previous scattering, the mutual shadowing calculation requires the knowledge of two previous scatterings (this will be clarified below). Formally this poses no problem as we merely reclassify the process as a Markov chain of order 2 (Takacs, 1960) instead of order 1. Computationally this requires that a given state of the process have an additional set of indices which defines its previous state; this correspondingly expands the size of the linear system and thus the order of the matrices which must be inverted for its solution. A second difficulty is that the shadowing is concentrated in a small angular range. For example, in the case of Saturn's rings this domain is only a few degrees (Kawata

and Irvine, 1975). This means that an adequate description of the effect requires the consideration of a large number of azimuthal Fourier components, each of which requires a matrix inversion in the method of the previous chapter. For economy it thus seems desirable to try a calculation intermediate between the classical computation (35) and the (formally possible) extension to all orders. This will give an estimate of the effort necessary to extend this formalism to all orders of scattering and of the relative magnitude of the effect of shadowing in the higher orders, as well as providing a check for the complete extension. It will be shown that a Markov chain of order 1 can remove the shortcomings of the classical method noted above. That is, the effect of shadowing will be accounted for in the second order and partially in all higher orders. Further, all orders will be internally consistent in that the higher orders may be calculated iteratively from the first two orders, which include the effects of mutual shadowing among the scatterers.

We consider the passage of a photon through the scattering layer as a stochastic process with each scattering a transition between possible states of the process with the same discretization of the atmosphere as in Chapter II. The following diagram (Figure 2) shows several consecutive transitions (scatterings) of a photon. It is scattered at A, then at B, and travels afterward in the direction Ω'. We must be able to calculate the probability of extinction at any point C along this path. In the absence of significant shadows, the probability of extinction depends only on the distance $\overline{BC}$. The effect of shadows is to correlate this probability with the previous path of the photon so that its

Figure 2. Most general scattering situation. Layers are the subdivision of the atmosphere used by the Markov chain formalism.


Upper boundary
A
B
C
D
Ω
α
Ω´
sub-
layers
Ground

likelihood of extinction becomes strongly dependent on the phase angle, α, between the two rays. The likelihood of extinction will be much less if this angle is small and the photon may travel close to its incoming path. This analysis shows that it is insufficient to know only B, but we must also know the location of point A if we are to determine the needed (transition) probabilities for a description of the photon's stochastic development after the scattering at B. Knowledge of point A allows calculation of the phase angle and also allows distinction between point D (where no shadowing correction applies) and point C which is close to the previous ray path AB. Therefore, in the presence of mutual shadowing, the computation of a transition probability at C depends in general on two previous scatterings, those at points A and B. This is the reason for the practical difficulty noted above in extending the calculation to include mutual shadowing in all orders of scattering.

However, the basic linear system of the Markov chain method may be written

$$I = \Pi_0 (1-Q)^{-1} R + \underset{\sim}{I}_s \quad , \tag{36}$$

where the total intensity I is found from the primary scattering vector $\underset{\sim}{I}_s$ and three probability matrices Π_0, Q, and R. The initial distribution vector, Π_0, represents the distribution of photons scattered exactly once in the atmosphere, which is taken to be the initial state of the Markov process. The scattering matrix, Q, contains the probabilities that a photon within the atmosphere be scattered but not escape. The matrix R gives the probability that a photon in a given state of the chain scatters and escapes.

The classical treatment for mutual shadowing would replace $\underset{\sim}{I}_s$ by the discretization of I_1^* as in (35). The inconsistency in the higher orders is readily apparent, as they are still determined linearly from Π_0. However, if we expand equation (36) in a Neumann series (this is always possible since the spectral radius of Q is less than unity) we have

$$I = \underset{\sim}{I}_s + \Pi_0 R + \Pi_0 QR + \Pi_0 Q^2 R + \ldots \tag{37}$$

Each term corresponds to an order of scattering, as in equation (34).

The most general scattering situation as illustrated in Figure 2 is described by the transition probabilities which are the entries of Q. However, the stochastic evolution of a photon upon either entering the scattering layer or just before leaving it is determined respectively by Π_0 and R. In addition, we see from (37) that the second order of scattering is given by the product $\Pi_0 R$. It will be shown that the entries of these two matrices can be computed to account for the effect of mutual shadowing consistent with a Markov chain of order 1. The Markov chain radiative transfer calculation will then give the second order intensity including shadowing. Further, since Π_0^* will be calculated including shadowing to be consistent with I_1^* and I_2^*, and all the higher terms may be found linearly from Π_0^*, the inconsistency of the series expansion (25) will be removed. Note, though, that because of the limitation to a chain of order 1, for the higher orders that effect of shadowing is only taken into account for the first and last scatterings that a photon suffers in the scattering layer.

The calculation of these probabilities is now the concern. To handle the azimuthal dependence of the radiation field, the calculation

is decomposed into a Fourier series. The matrix algebra in equation (36) is then performed for each Fourier component, and finally an inverse transformation yields the total intensity as in equation (5) in the previous chapter. As above, the effect of mutual shadowing will be to decrease the probability of extinction along a ray path of optical length τ from $\exp(-\tau)$ to $\exp(-\tau + n_0 V')$, where $n_0 V'$ is given by (27) or a similar calculation.

The transition probability that a photon last scattered downward in direction μ_j (the cosine of the j-th element of finite set of zenith angles) at a depth τ' in the layer will scatter and escape through the upper boundary of the layer in a direction μ_e is given by

$$R(\mu_j,\tau';\mu_e) = e^{-\tau'/\mu_e} R^1(\mu_j,\mu_e,\tau_0-\tau') \quad . \tag{38}$$

Here $R^1(\mu_j,\mu_e,\tau)$ is the singly scattered reflection from a homogeneous scattering layer of optical depth τ, and τ_0 is the total optical depth of the layer. By "homogeneous" we mean that the average properties of the atmosphere are independent of optical depth and not the more restrictive sense that $g(\vec{r})$ be unity. The transition probability for the Markov chain $R_{\ell e}$, which we shall explicitly write as $R(j,n;e)$, is the same as R in (38) except that the photon is constrained to have its last scattering at $\tau' \in [\tau_n,\tau_{n+1}]$, i.e., within the n-th sub-layer in the subdivision of the atmosphere. This will be given by

$$R^*(j,n;e) = R^*(\mu_j,\tau_n;\mu_e) = e^{-\tau_n/\mu_e} R^{1*}(\mu_j,\mu_e,\tau_0-\tau_n) \quad . \tag{39}$$

In the Markov chain without shadowing, this probability is given by the average of $R(\mu_j,\tau';\mu_e)$ over the sub-layer

$$R(j,n;e) = \int_{\tau_n}^{\tau_{n+1}} \frac{d\tau'}{\Delta\tau_n} R(\mu,\tau';\mu_e) \quad , \tag{40}$$

where $\Delta\tau_n$ is the optical depth of sub-layer n. However, those integrals could be determined analytically in the absence of shadowing. Since the integrals (40) must be in the present case be computed numerically, the benefit of averaging over the layer n is outweighed by the increase of computation required to perform the average. Instead, we evaluate the integrand in (40) at the upper boundary (τ_n) to give the transition probability (30). Thus, we assure that we overestimate the effect of shadowing on these transition probabilities. As is evident, shadowing has no effect on the transition probabilities in R when μ_j is upwelling.

The vector Π_o is the initial distribution for the Markov chain, which in the present formalism is the distribution of photons scattered once. To assure consistency of the higher orders of scattering with the first order, probability must be conserved. This is accomplished as follows. Let E be the diagonal matrix whose entries are the probability of extinction within the atmosphere for a photon occupying that state. For conservative scattering the row sums of Q are the entries of E. We require (recalling that asterisks denote inclusion of mutual shadowing)

$$\underset{\sim}{I}_s^* = \Pi_o^* (1-E) \quad . \tag{41}$$

We can satisfy this consistency requirement by setting

$$\frac{\Pi_o^*}{\Pi_o} = \frac{\underset{\sim}{I}_s^*}{\underset{\sim}{I}_s} \quad , \tag{42}$$

since

$$\underset{\sim}{I}_s = \Pi_o(1-E) \quad .$$

This gives for the m-th Fourier component

$$\pi_o^{(m)*} = \Pi_o^{(m)} \int_o^{\pi} \frac{d\phi}{\pi} \cos m\phi \, \Phi(\alpha) \, \exp[n_0 V']/\Phi^{(m)}(\mu_0,\mu_i) \quad , \qquad (43)$$

where $\Phi^{(m)}$ is the m-th Fourier component of the phase function, and μ_0 and μ_i are respectively the zenith angle cosines for illumination and scattering. As with the calculation of R^*, we overestimate the effect of shadowing by selecting τ_{n+1} (the bottom boundary of layer n) as the characteristic optical depth for this calculation. These corrections ignore higher order effects having a Markovian dependence of order 2. As stated above, these would require increasing the matix size and also modifying the elements of Q, but would include the general situation in Figure 2.

A Markov chain radiative transfer code as described in Chapter II was modified to calculate these new matrices by numerical quadrature of equations (39) and (43) assuming spherical, non-interacting particles with fractional volume D. With D=0, this code was compared with a standard doubling routine (Hansen and Travis, 1974). The error (difference) was less than 1% and could be mostly attributed to removing the averaging in the calculation of R^* (equation 40). Other errors that were introduced in modifying the formalism were smaller by a factor of ten.

To remove this (D=0) zero-point error, the calculations for non-zero density were compared with the modified code with D=0. Therefore, the calculated results are the difference in intensity caused by mutual

shadowing, which should likewise be accurate to within a few per cent. This error could be made smaller by subdividing the atmosphere more finely, with a concurrent increase in the computer time required.

The magnitude of the effect of mutual shadowing is determined by the physical parameters of the scattering layer. The effect is not strongly influenced by single scattering albedo and total optical depth. Since $\tilde{\omega}_0$ is only a constant outside the integration, its effect scales as $(\tilde{\omega}_0)^n$, which for these calculations is $(\tilde{\omega}_0)^2$. Schoenberg (1929) has also shown that the effect of mutual shadowing is relatively independent of τ_0 for atmospheres that are optically thick.

Since the light scattering in Saturn's rings seems a natural application for these calculations, the parameters for the following computations were chosen in a range that seems plausible for Saturn's brighter, B ring. For all these cases, the albedo for single scattering was taken as 0.85 and the total optical depth as 1.0. The fractional volume occupied by the scatterers was either D = 0.0, D = 0.0125, or D = 0.1. An important physical parameter is the phase function for scattering. Four distinct cases were considered, modeled by the Henyey-Greenstein function,

$$P(g,\alpha) = (1-g^2)/(1 + g^2 - 2g \cos \alpha)^{3/2} \quad . \qquad (44)$$

These were

1) strongly backscattering $g = -0.7$

2) strongly forward scattering $g = +0.7$

3) Isotropic scattering $g = 0.0$

4) Forward and backward scattering:

$$P(\alpha) = b\, P(g_1,\alpha) + (1-b)\, P(g_2,\alpha)$$

$$\text{with } g_1 = -0.5, \quad g_2 = 0.5 \quad b = 0.404 \quad .$$

This last case was found to be a likely phase function for the individual particles in Saturn's rings from an analysis by Esposito and Lumme (1977).

For each set of parameters considered, the effect of mutual shadowing in the diffuse reflection was calculated three times: once including shadowing only in $\underset{\sim}{I}_s^*$, then including shadowing in $\underset{\sim}{I}_s^*$ and R^*, and finally in $\underset{\sim}{I}_s^*$, R^*, and Π_o^*. This allows us to separate the effect due only to shadowing in the multiple scattering, and to compare the effect of Π_o^* and R^* on the observed effect.

The effect of inclusion of shadowing in the higher orders of scattering will be measured by the quantity

$$X = [I(\alpha=0.1°) - I(\alpha=6.0°)] - [I_1(\alpha=0.1°) - I_1(\alpha=6.0)] \quad , \qquad (45)$$

where I is the total diffuse intensity and I_1 the intensity due to single scattering. This gives a useful indication of the rise of the phase curve near $\alpha=0°$ caused by <u>multiple</u> scattering. We can refer to this as the multiple scattering contribution to the "opposition effect". Table 2 shows how the quantity is increased by inclusion of mutual shadowing.

As can be seen by comparing the second and third columns and fourth and fifth columns with those in the first column (where shadowing is not included), the increase in the opposition effect arises about equally from including shadowing in Π_o^* and R^*. The effect increases with D and $|g|$, as could be expected, and always serves to sharpen the phase curve near $\alpha = 0°$. However, for the range of parameters, the effect is much too small to be observable in any astronomical object, always less than 10^{-3} of the total intensity at the relevant angles. This is much

Table 2

X, Opposition effect due to multiple scattering. See equations 44 and 45. Calculations for solar flux F=1 and zenith angles $\mu = \mu_0 = 0.44$. For comparison, the opposition effect due to single scattering lies in the range 0.01 (forward scattering) to 0.2 (backscattering). Negative values indicate decreased opposition effect.

Phase Function	Without shadowing Density = 0.0	Density = 0.0125 Shadows in R^*	Density = 0.0125 Shadows in R^* & Π_o^*	Density = 0.1 Shadows in R^*	Density = 0.1 Shadows in R^* & Π_o^*
Backscattering	9.2×10^{-4}	9.4×10^{-4}	9.7×10^{-4}	10.3×10^{-4}	13.7×10^{-4}
Forward scattering	-1.6×10^{-4}	-1.2×10^{-4}	-1.1×10^{-4}	-1.0×10^{-4}	-0.9×10^{-4}
Isotropic	0.0	0.0	0.0	0.0	0.0
Saturn's rings	3.8×10^{-4}	3.9×10^{-4}	4.1×10^{-4}	4.2×10^{-4}	4.9×10^{-4}

smaller than some other improvements to the classical theory which only involve single scattering: the effect of dispersion in particle radius (Bobrov, 1961) and the effect of the finite angular size of the solar disk (Kawata and Irvine, 1975). Note that the effect can be much larger than that predicted by Irvine (1966) if the phase function is highly anisotropic.

In Table 3, the shadowing effect in the second order is decomposed into its Fourier components for a particular case found applicable to Saturn's rings (Esposito and Lumme, 1977). As can be seen, mutual shadowing has a large effect on the higher components of the radiation field, especially if D is large. This can sometimes increase the magnitude of an individual Fourier component by a factor of 150 over the non-shadowing case. However, in the standard problem these higher components contribute very little to the observed scattered light. This might not be true in a highly structured scattering layer like a planetary regolith. In fact, the structure may be highly anisotropic so that the higher Fourier components may be actually enhanced by multiple reflections. In such a case, the inclusion of higher orders of mutual shadowing is clearly essential. Unfortunately, the present formulation is not necessarily helpful in this regard because such a physical situation may violate the assumptions that $D^2 << 1$ and that the particles are not subject to strong many-body interactions. Further, it is not clear in such a rough surface that the radiative transfer excluding shadowing is well modeled by the standard formulation. On the other hand, experimental studies of rough surfaces show that their light scattering may be empirically modeled by the classical mutual shadowing theory (Veverka, et al.,

Table 3

Per cent change due to higher order mutual shadowing in the value of the Fourier component of the total higher order scattering. Calculations are for solar flux F=1 and zenith angles $\mu = \mu_0 = 0.44$. The contribution of first order scattering is ~ 0.5 which may be compared with column 2.

Fourier Component	Total Higher Order Scattering	Per Cent Change	
	Density = 0.0	Density = 0.0125	Density = 0.1
1	0.1	0.4	3
2	-0.003	3	21
3	0.005	2	11
4	-0.0003	7	40
5	0.0002	4	35
6	-2×10^{-5}	18	110
7	1×10^{-5}	20	110
8	-1×10^{-6}	50	300
9	6×10^{-7}	65	380
10	-1×10^{-7}	190	980
11	2×10^{-8}	230	1400
12	-5×10^{-9}	700	3600
13	1×10^{-9}	1000	5300
14	-3×10^{-10}	2300	11000
15	1×10^{-10}	3000	15000
16	-2×10^{-11}		

1977). This may mean that the extensions developed above are also relevant to such a surface. However, it appears that further theoretical work is necessary before we can compute the photometric properties of regolith-like surfaces.

Thus, while the Markov chain method allows inclusion of the effect of mutual shadowing to higher orders, and removes some inconsistency of the classical method, calculation shows that the effect is negligible, at least for Saturn's rings. Since the magnitude of the effect decreases so greatly from the first to the second order of scattering, an extension to all orders of scattering (which is formally possible) does not seem worthwhile at this time. Likewise, previous calculations which considered only primary scattering are shown to be quite accurate.

CHAPTER IV

PHASE CURVES OF SATURN'S RINGS AND THEIR ANALYSIS

Introduction

The most distinctive characteristic of the photometric behavior of Saturn's rings is the variation of their brightness with solar phase. The rings brighten appreciably as Saturn approaches opposition; this increase is much larger than could be expected from extrapolation of their phase curves at larger phase angles. This is the "opposition effect", which has been explained for the classical model of the rings as being due to the mutual shadowing among the scattering particles (see Chapter III) at least in a layer many particles thick (e.g. Pollack, 1975). Studies of this phase variation yield the opportunity to determine the characteristics of the individual particles as scatterers and shadow-casters. This can be done if we can formulate consistent models of the rings which give rise to the photometric variation that is observed.

Since Müller (1893) first reported an anomalous brightening for Saturn's rings, numerous studies of the phase curve for Saturn's rings have been made. The observations themselves are somewhat inhomogeneous, as noted by Franklin and Cook (1965). Most of these observations did not distinguish between the two brighter rings of Saturn (rings A and B), although this is straightforward in good quality photographic images. Recent observations are reported by Franklin and Cook (1965), Bobrov (1970), Irvine and Lane (1973), and Lumme and Irvine (1976). Of these,

only Lumme and Irvine consider the A and B rings separately, and their data are limited by a lack of observations near opposition. The phase curves of Franklin and Cook may be taken as standard because the scatter of individual measurements is quite small and the curves themselves are consistent with the other reported observations. Franklin and Cook measured the entire Saturn system photoelectrically and were able to subtract the light due to Saturn's disk by comparison with contemporaneous photographic images. It is worthwhile to note that at the level of accuracy of their observations, no difference was found between the two rings in the shape of their phase curves.

Lumme and Irvine (1976) made an improvement in analysis of photographic photometry of Saturn's rings by defining a method to determine the absolute brightness of the rings by comparison with the center of Saturn's disk. The absolute calibration was based on scans of photographic images of Saturn and the geometric albedo of the entire planetary disk measured by Irvine and Lane (1971). Although there are considerable uncertainties in the absolute scale so defined (see Esposito and Lumme, 1977), this procedure allows photographic data to be analyzed in an absolute sense to compare with radiative transfer models of the rings. These models must give a satisfactory fit not only to the shape but also to the scale of the phase curve.

The analysis of the phase curves for Saturn's rings has predominantly interpreted the opposition peak as the result of mutual shadowing among the scatterers. This explanation was first advanced by Seeliger (1887), who also made the first quantitative analysis of the effect of mutual shadowing (see Chapter III). Improvements to this classical, geometrical

optics method for the explanation of the ring phase curves have been made to include a dispersion in particle size (Bobrov, 1961), diffraction into the shadow (Franklin and Cook, 1965), the finite size of the sun and solar limb darkening (Kawata and Irvine, 1975). All of these latter authors also included the effect of multiple scattering in the rings. Recently some doubt has arisen about the applicability of a model for the rings that is many particles thick. This is based on the dynamical expectation that due to inelastic collisions among the particles the rings will have relaxed to a layer roughly one particle thick. In such a case the opposition effect must be explained as being due to the reflection properties of the individual particles alone (Price, 1974, 1977). The calculation of the dynamics of the ring particles is, however, not an easy one, and some skepticism must be placed on the monolayer prediction in the light of recent computations (Brahic, 1977), so that the case is not yet closed.

An observational study of the rings of Saturn may hope to answer a number of questions: 1) are the two rings different, 2) do the phase curves show some dependence on color, and 3) how does the phase curve depend on the declination of the earth and sun relative to the ring plane ("tilt" angle)? We note that the majority of the existing observations do not distinguish the rings; there is no complete phase curve for the rings in the red; and there is a serious lack of data on the phase curves except near maximum declination of the earth and sun. This study will provide some data on all three questions.

The analysis and interpretation of the observations will also have well-defined goals. The first is to determine the parameters of a multiple scattering model which reproduces the phase curves. The second

is to compare these parameters with those determined in previous studies, not only of the phase curves but also of other photometric behavior of the rings. An example is the variation of brightness with changing tilt angle, which was studied by Price (1973) and Esposito and Lumme (1977). Also, Kawata and Irvine (1975) have studied the infra-red emission of the rings, which also constrains the optical parameters of a ring model. Another goal is to determine if the differences between the rings necessitate differences in the individual particle properties, or (as found by Lumme and Irvine (1976) and Esposito and Lumme (1977)), whether these can be explained as differences in the thickness and density of the rings. The radiative transfer calculations will be carried out by a Markov chain formalism (Chapter II), which includes shadowing in the primary scattering only (Chapter III).

Observations and Reductions

The observations for this study consist of photographs obtained in January and February 1977, made at four different observatories to ensure relatively uninterrupted coverage of high quality. The four telescopes had identical optical systems of equal focal length. Cameras, color filters and emulsions were also identical. All of the films were calibrated and processed at Lowell Observatory using completely standardized procedures.

The telescopes were 61-cm f/75 Cassegrains of the International Planetary Patrol network (Baum, 1973), located at Mauna Kea Observatory in Hawaii, Perth in Australia, Cerro Tololo Inter-American Observatory in Chile and Lowell Observatory in Arizona. The Patrol cameras are

automatically sequenced 35-mm cinecameras that have been specially designed to record the place, date, time and filter on each frame. They are also equipped to provide sensitometric calibration at the telescope.

In order to avoid the slight non-uniformities of transmission typical of evaporated coatings, field-viewing beam splitters were removed from the Patrol cameras, and absorption-type color filters were substituted for the interference filters ordinarily used. In addition, the entire telescope was inverted each night so that images were photographed in two orientations 180° apart to detect any asymmetry in the telescope itself. The composition of the filters, effective wavelengths (including emulsion response), band widths (FWHM) and typical exposure times are listed in Table 4. The emulsion was Kodak 2498 RAR processed to a gamma of 1.25 in a GAF Transflo 1207 film processor.

A total of 25,000 acceptable images were obtained from these observations. From these, 257 single images were selected on the basis of image quality and distribution in phase angle and color. The distribution of images was the following: red, 80; green, 71; blue, 53; and UV, 42. These images were scanned using a Boller and Chivens PDS 1010A image-digitizing microdensitometer coupled to a PDP-11 computer at the Planetary Research Center at Lowell Observatory. This yielded a digital representation of the images that was stored on magnetic tape. Each picture element was the average density in a 50μ square of the original image with a possible range of 1024 gray levels. The size of a picture element corresponds to an angular distance of 0".225 on the sky. The sensitometric exposures of the calibration wedges were digitized at the same time.

These digitized images were further reduced at the Jet Propulsion

Table 4

Color	Schott filter glass numbers	Effective wavelength	Bandwidth (FWHM)	Typical exposures
Red	OG550 (3mm)	5900Å	730Å	5 sec
Green	BG23(3mm) + GG7(3mm)	5350	940	12
Blue	BG12(1mm) + GG13(2mm)	4150	720	14
UV	UG1 (1mm)	3560	690	50

Laboratory in Pasadena, California. At their Image Processing Laboratory the images of the calibration wedges were read to determine the average density of exposure in each calibration spot. Since the transmission of each wedge is known, a smooth curve may be fitted to these average density numbers to determine the specific intensity corresponding to any gray level as a function of its density. The intensity (relative to some arbitrary normalization) was modeled by the function

$$I = \exp[A_0 + A_1\sqrt{D} + A_2 D] \quad , \tag{46}$$

where D is the density of exposure. The coefficients A_0, A_1, A_2 are determined from the nightly calibration exposure by a least square fit between the known transmission of each wedge and the darkness of its exposure on the film. This is done for the six density wedges which bracket the normal exposures for the rings. Although the calibration exposure is not taken through the same filters as the Saturn observations, the variation in wavelength response of the film is not large enough to cause a difference of more than a few per cent in the photometry of the rings relative to each other and the disk of Saturn. This is supported by a comparison of the ring A/ring B brightness ratio for a selection of these images (Lumme, et al, 1977) with those from Lumme and Reitsema (1977), for which the calibration exposures were taken through the observing filters. Lumme (1977, private communication) finds no significant difference to the accuracy of the data.

In photographic photometry a natural reference point for brightness measurements of the rings is the center of Saturn's disk. The data presented here refer to the brightest point in each ring compared to the

center of a smoothed scan of Saturn's disk along the major axis. Because no systematic east-west difference is found, a measurement is the average of the brightness at the two ansae. To determine the photometric variation of the rings we need to know both the variations of the brightness of the center of disk and its absolute brightness.

Our treatment in this regard will follow Lumme and Irvine (1976). They note an absence of reliable data on possible short-term and long-term temporal variations of the disk of Saturn. The variations with phase angle and latitude of the center are somewhat better understood. In blue and visual (green) the phase curve for the disk is known from Franklin and Cook's (1965) data. From study of microdensitometer scans of green images along the minor axis at various latitudes, Lumme and Irvine (1976) have determined the limb darkening law for Saturn and modeled the latitudinal variation of brightness as a uniformly bright equatorial zone superimposed on an otherwise uniform disk. In combination with the monochromatic albedo of Saturn measured by Irvine and Lane (1971) this gives the absolute brightness at the center of the disk. For the declination of the earth and sun at the date of these observations ($B \sim B' \cong 17°$) this yields for green images

$$I_D/F = 0.55 \pm 0.1 \quad . \tag{47}$$

I_D is the specific intensity of the disk center and πF is the incident solar flux at Saturn. This will be used later to find the absolute phase variation of the rings, at least in green. Aside from the correction for the phase variation of the disk in blue and green and the absolute normalization for green, all observations were also adjusted to

a "mean" tilt angle, $\bar{B}$, so that all observations correspond to equal declination of the earth and sun $B = B' = \bar{B} = \arcsin 0.288 = 16.8°$. This correction is applied by multiplying each observed brightness by the factor $0.288(1 + \frac{\sin B}{\sin B'})/(2 \sin B')$. Because the data is so near to opposition, this correction is always less than 3%. Further, using realistic models of the rings Esposito and Lumme (1977) have determined the accuracy of the extrapolation from unequal angles, B, B', to a mean tilt angle $\bar{B} = \arcsin [2 \sin B \sin B'/(\sin B + \sin B')]$, which is the major part of this correction, to be better than 0.1%.

Phase Curves

Figures 3-6 present the observed phase curves in each of the four colors after the above corrections. The range of the observations is from a phase angle of 3.6° before opposition to 0.12° on opposition night and to 1.65° after opposition. The ordinate of the figures is the brightness in magnitudes relative to that of the brightest part of the B ring on opposition night. The open circles are the observations from Cerro Tololo (CT), the filled circles those from Mauna Kea (MK) before opposition, and the crosses are the observations after opposition (all from Mauna Kea). No images from Lowell or Perth were judged to be of high enough quality for inclusion. For some colors a significant difference of up to 15% occurs between the data from Cerro Tololo and Mauna Kea, showing that despite the effort expended, exact uniformity has not been achieved. The discrepancy may be partially due to systematic differences in fitting the calibration curves (equation 46) to the sensitometric data from the two observatories, but in any case sets a lower

Figure 3. Red phase curves for Saturn's rings relative to Saturn's disk.

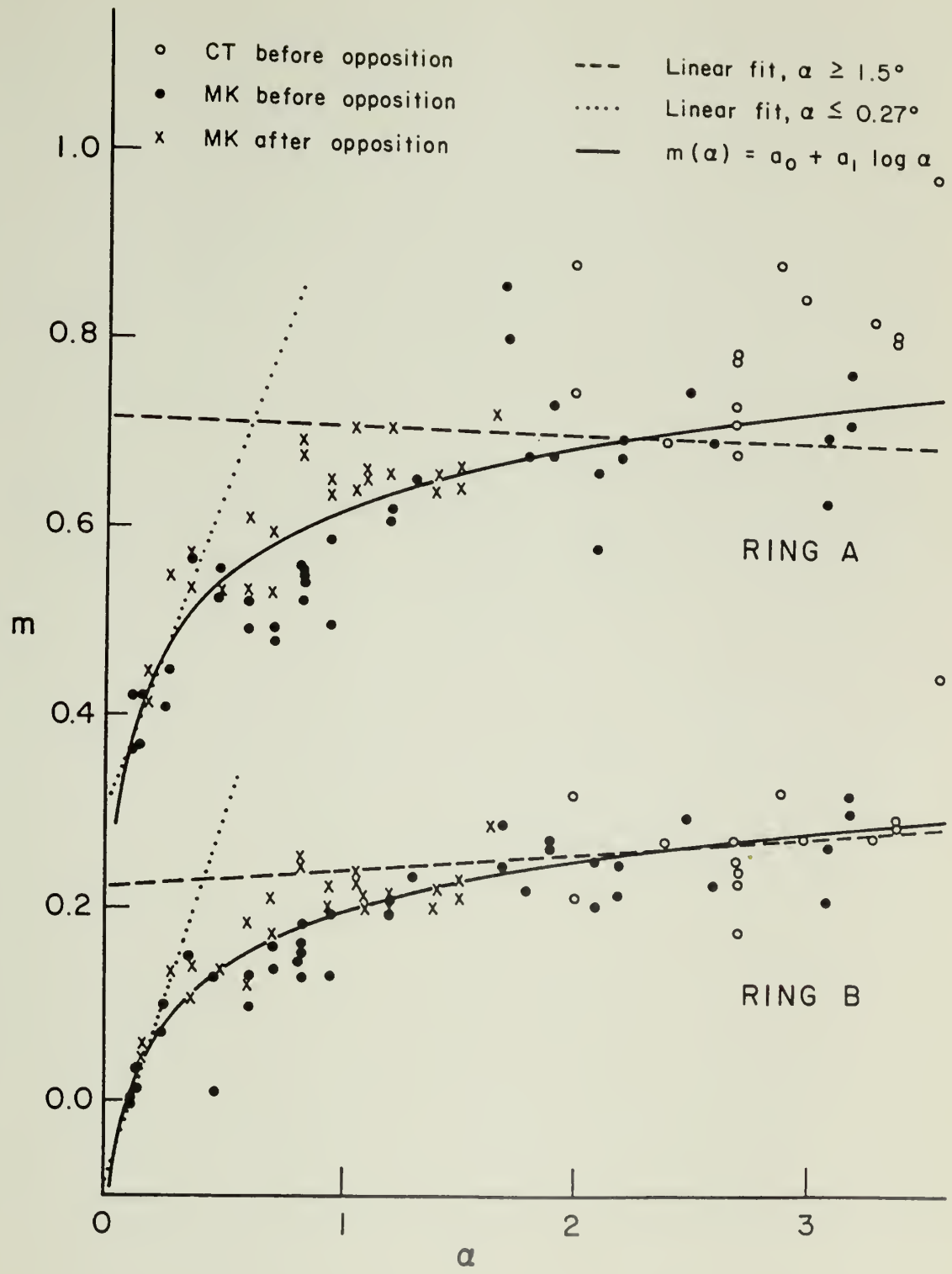

CT before opposition
MK before opposition
MK after opposition
Linear fit, α ≥ 1.5°
Linear fit, α ≤ 0.27°
m(α) = a0 + a1 log α
RING A
RING B
m
α
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3

Figure 4. Green phase curves for Saturn's rings. These are corrected for the phase variation of the disk. Symbols as in Figure 3.

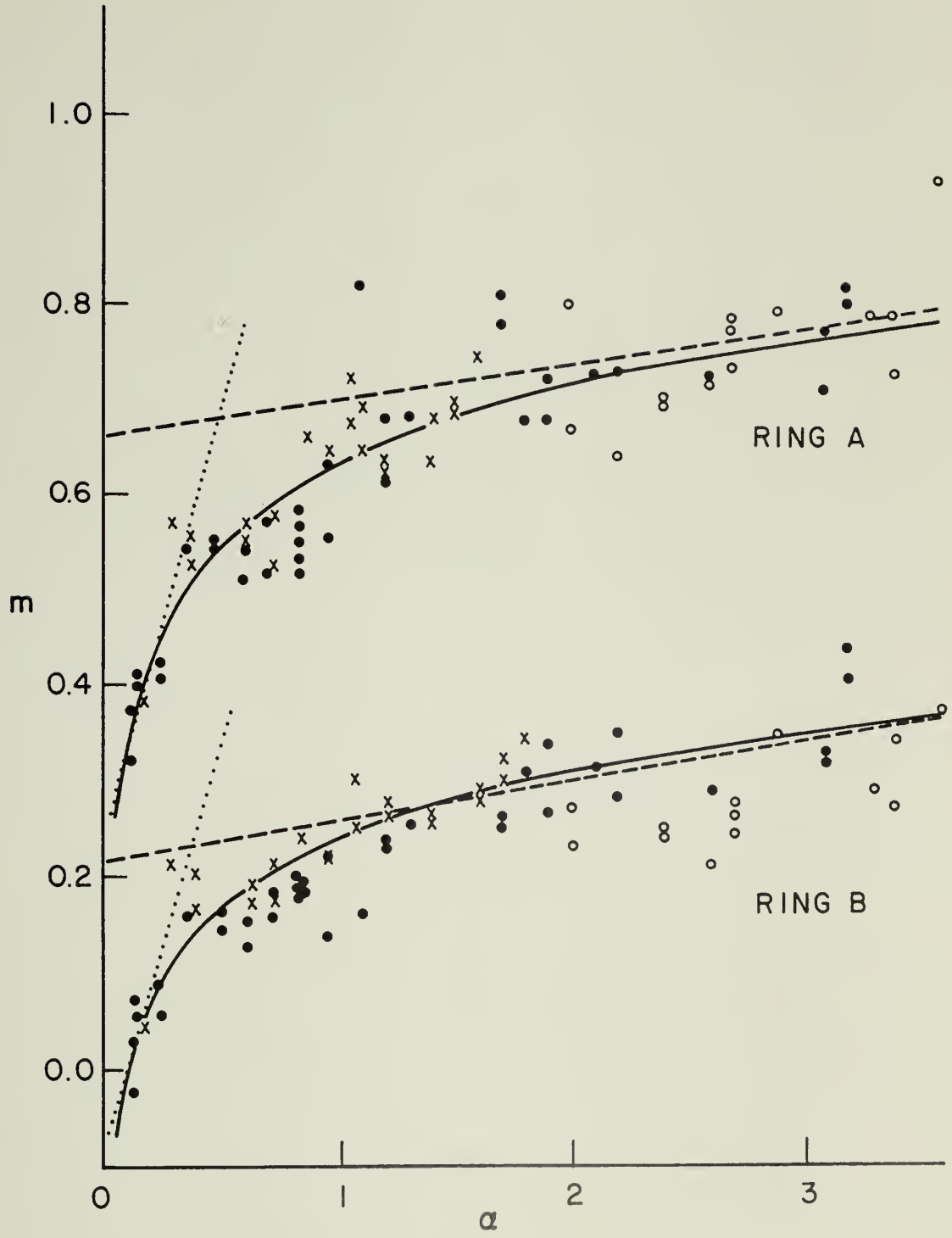

1.0
0.8
0.6
0.4
0.2
0.0
m
RING A
RING B
0
1
2
3
α

Figure 5. Blue phase curves for Saturn's rings corrected for the phase variation of the disk. Symbols as in Figure 3.

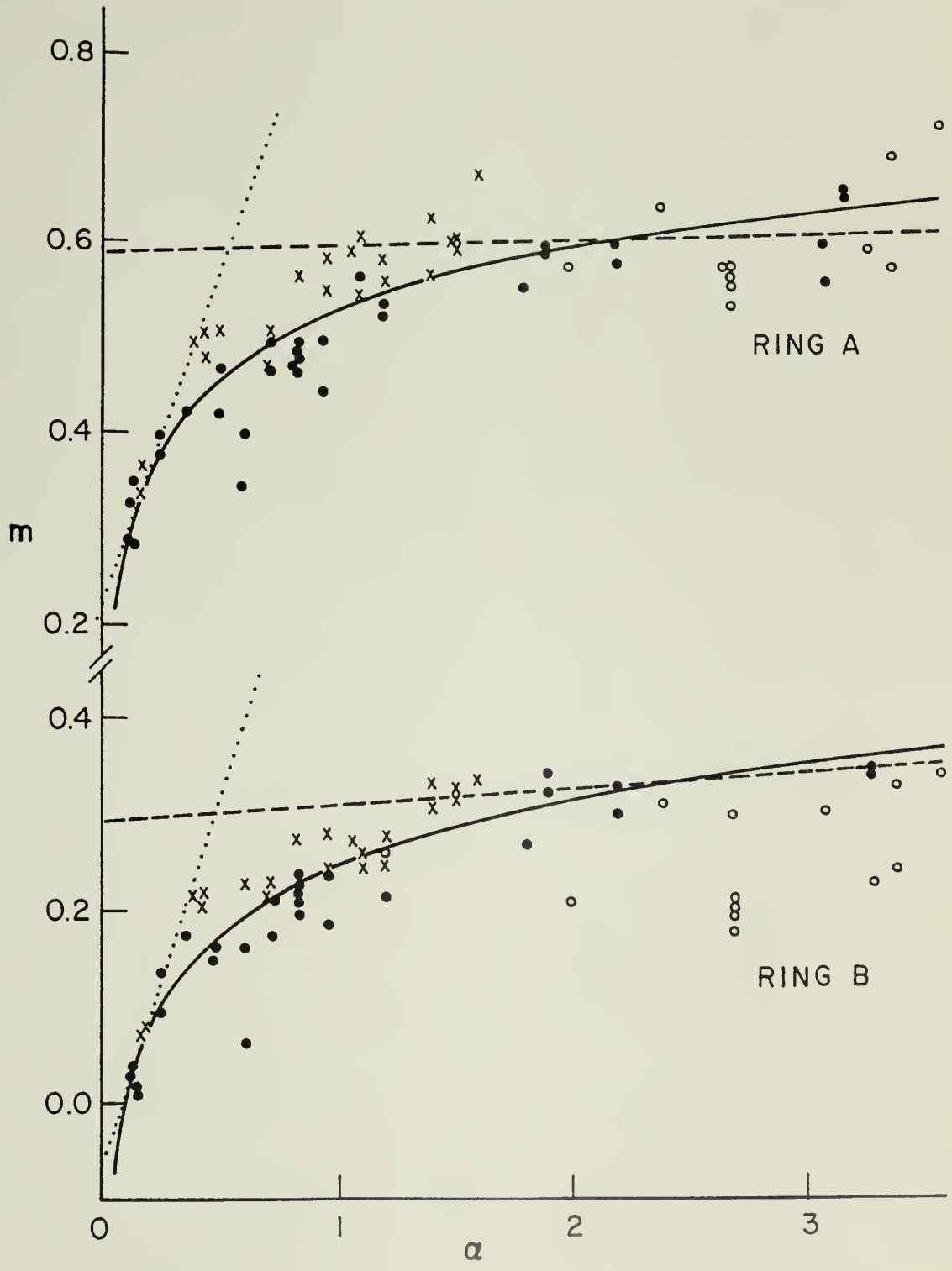
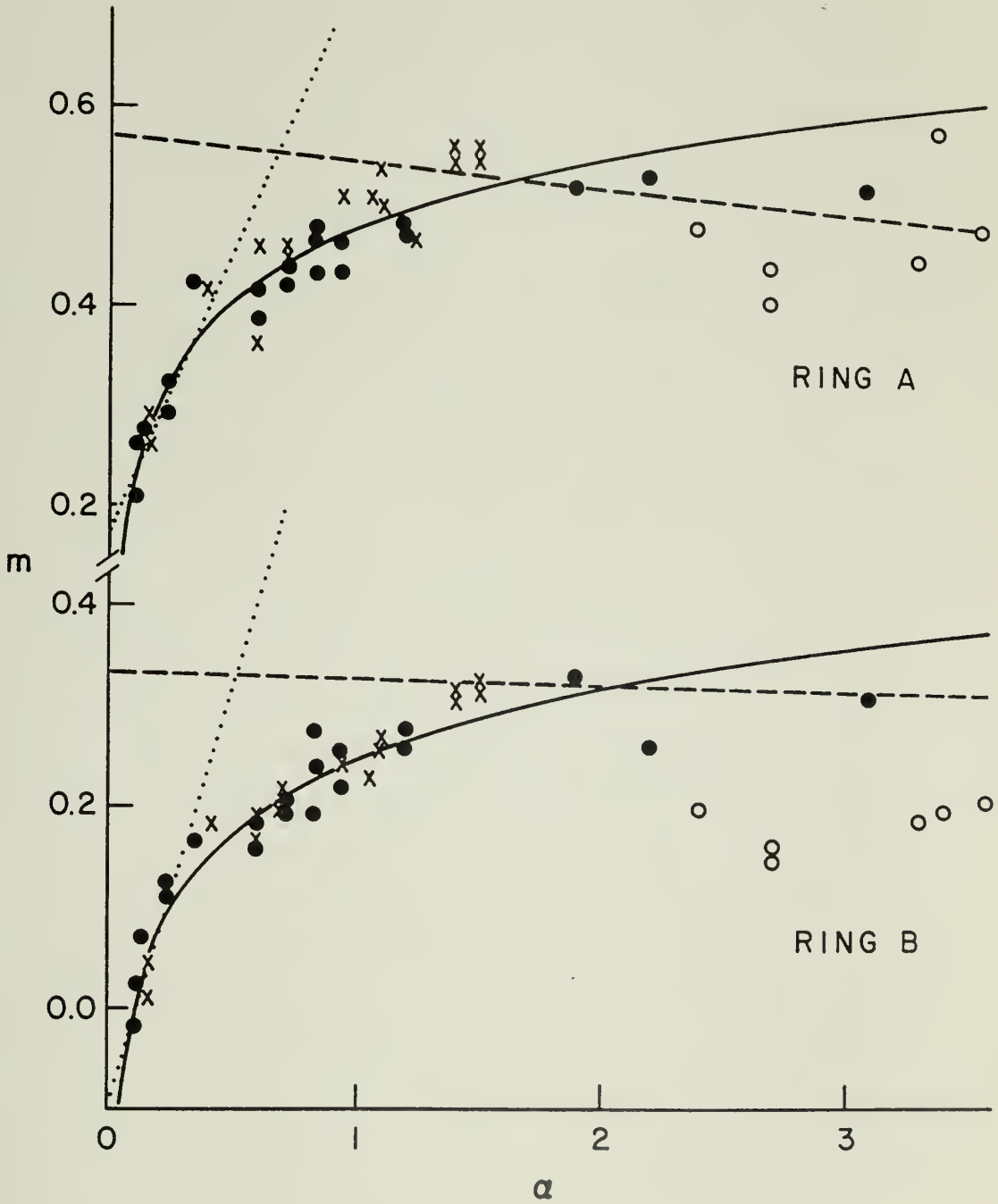

0.8
0.6
0.4
m
0.2
0.4
0.2
0.0
RING A
RING B
0
1
2
3
α

Figure 6. UV phase curves for Saturn's rings relative to Saturn's disk. Symbols as in Figure 3.


0.6
0.4
0.2
m
0.4
0.2
0.0
RING A
RING B
0
1
2
3
α

limit to the accuracy of these data.[1] The figures also show several least squares fit to the data, which will be discussed later. All of these fits were made to the Mauna Kea data alone, first, because the majority of the high quality observations are from Mauna Kea and, second, because the inclusion of the Cerro Tololo data substantially increased the standard errors, by as much as a factor of two in some cases. It is important to realize that because of this systematic error (and perhaps others undetected) the actual uncertainty in any conclusions may be substantially greater than the computed standard errors.

Figures 7 and 8 give the brightness ratio I_A/I_B of ring A to ring B. In all four colors this ratio appears to be independent of phase angle, as was found in B and V by Franklin and Cook (1965) when the tilt angle was $\bar{B} \sim 26°$. For the present observations $\bar{B} \sim 17°$. In red and green, however, there may be some indication that this ratio is slightly higher near opposition, indicating a stronger opposition effect in ring A. Although no significant difference is seen in these figures, it is possible that a different method of analysis may show a real difference in the phase curves. By comparing fits to various portions of the phase curves for each of the rings this difference may be accentuated, as will be shown later. Note that the ratio increases with decreasing wavelength. This is consistent with an increased contribution for ring B due to multiple scattering where the albedo is greatest at longer wavelengths (Irvine and Lane, 1973).

The magnitude of the opposition effect Δm will be defined as the difference in the intercepts of a linear fit to those observations with $\alpha \geq 1.5°$ and a linear fit to observations with $\alpha \leq 0.27°$. This was the

Figure 7. Brightness ratio, ring A to ring B. Red (top), green (bottom). Symbols as in Figure 3.

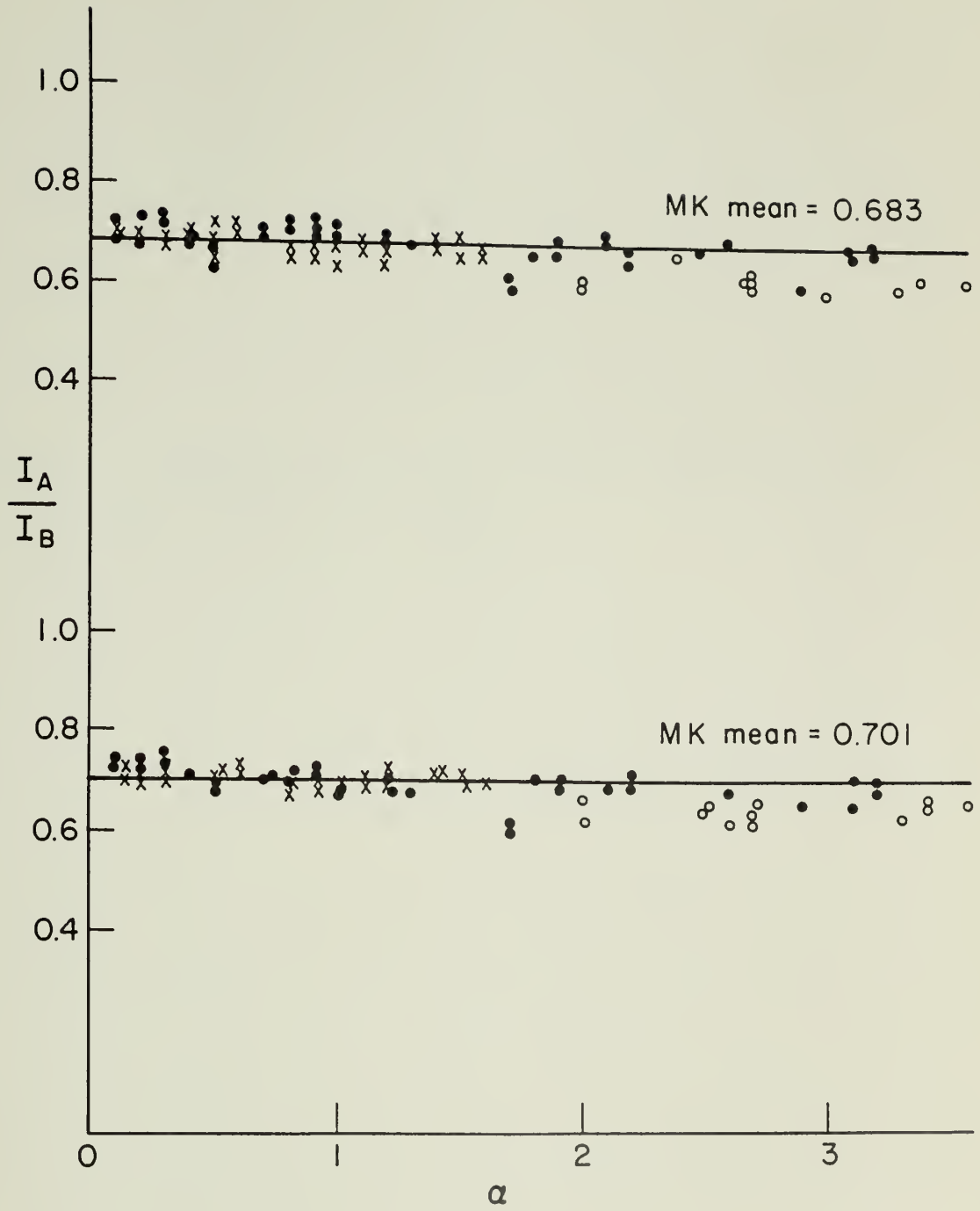
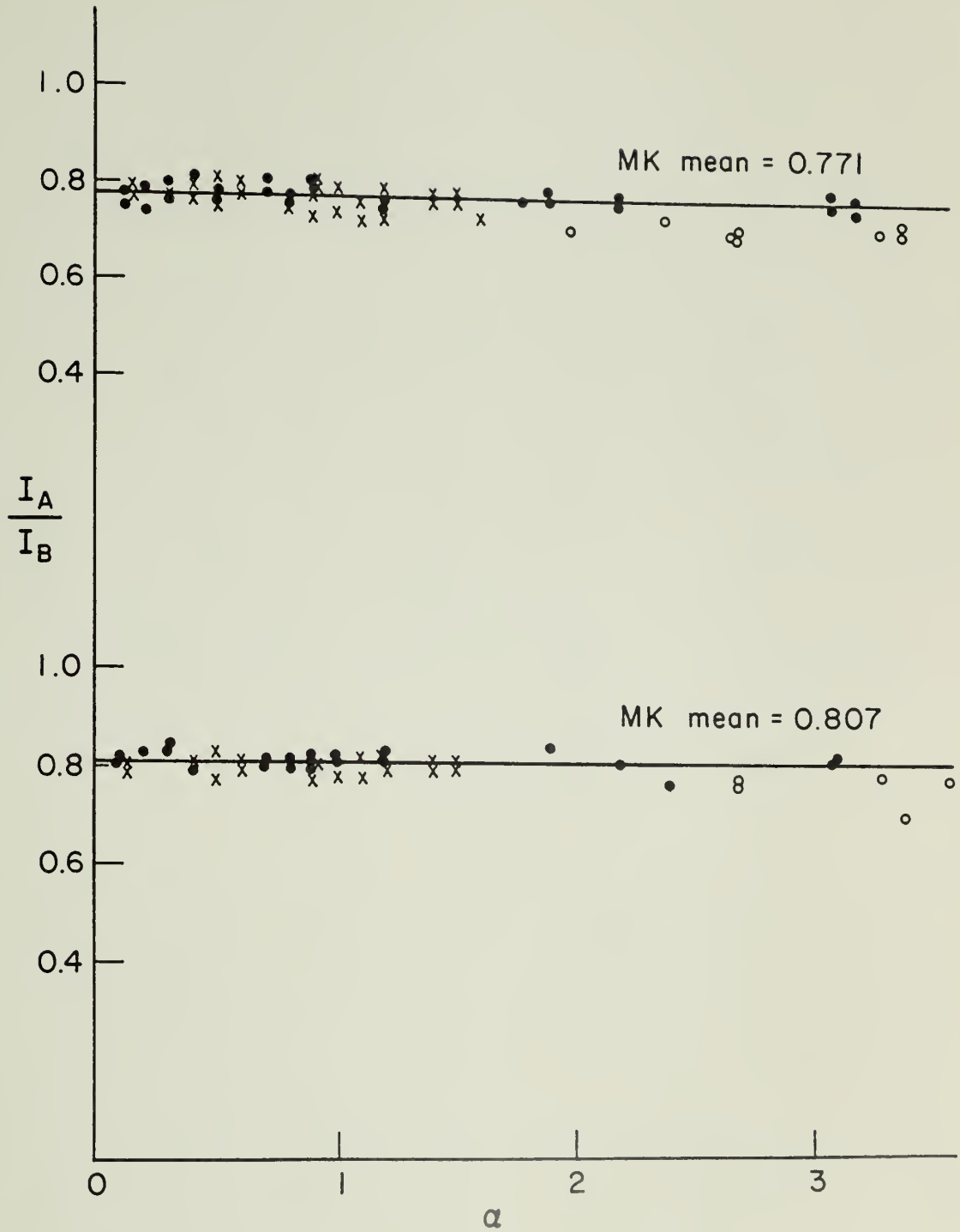

1.0
0.8
0.6
0.4
$\frac{I_A}{I_B}$
MK mean = 0.683
1.0
0.8
0.6
0.4
MK mean = 0.701
0
1
2
3
α

Figure 8. Brightness ratio, ring A to ring B. Blue (top), UV (bottom). Symbols as in Figure 3.


1.0
0.8
0.6
0.4
$\frac{I_A}{I_B}$
MK mean = 0.771
1.0
0.8
0.6
0.4
MK mean = 0.807
0
1
2
3
α

method used by Lumme and Irvine (1976) to analyze Franklin and Cook's data for both rings together. It differs somewhat from the method of Irvine and Lane (1973), who used a quadratic fit to those points with $\alpha \leq 1.5°$ to compare with the linear extrapolation from $\alpha \geq 1.5°$. These fits appear in Figures 3-6 and Δm is tabulated in Table 5. Since the phase curve is nearly linear at larger phase angles, we may take the slope of the line for $\alpha \geq 1.5°$ as the phase coefficient for the ring, $k = \frac{dm}{d\alpha}$, excluding the opposition surge. These also appear in Table 5. The phase coefficients are very poorly determined because of the sparseness of data in the region $\alpha \geq 1.5°$. Because of the large standard errors they are consistent with previous determination of the phase coefficients by Lumme and Irvine (1976) for $\bar{B} \sim 26°$ and Lumme and Reitsema (1977) for $\bar{B} \sim 18°$, but are otherwise not too informative.

In agreement with Lumme and Irvine (1976) and Bobrov (1970) we found that the entire phase curve may be well fitted by the function

$$m(\alpha) = a_0 + a_1 \log \alpha \quad . \tag{48}$$

This was true for both rings in every color. Since the scale in magnitudes is arbitrary, only the parameter a_1 is physically relevant: it gives the shape of the phase curve in the entire range $0.12° \leq \alpha \leq 3.6°$. For these observations the residual variance was quite small, and in all colors the fit was very good, as can be seen in the figures. Table 6 gives the parameter a_1 along with the results from the observations of Lumme and Irvine (1976) and Franklin and Cook (1965), which were analyzed in this way by Lumme and Irvine (1976). In all cases, the phase angle α is measured in degrees. In comparison with the photographic observations

Table 5

Opposition effect and phase coefficients for Saturn's rings. Uncertainties are standard deviations. Mauna Kea data only.

	Δm, opposition effect		k, phase coefficient[a]	
Color	Ring A	Ring B	Ring A	Ring B
Red	0.41 ± 0.08	0.30 ± 0.04	-0.01 ± 0.03	0.02 ± 0.02
Green	0.41 ± 0.08	0.30 ± 0.07	0.04 ± 0.02	0.05 ± 0.02
Blue	0.36 ± 0.05	0.36 ± 0.04	0.01 ± 0.02	0.02 ± 0.02
UV	0.39 ± 0.04	0.43 ± 0.05	-0.01 ± 0.01	-0.01 ± 0.01

[a]In green and blue, this includes a correction for the phase coefficient of the disk. In red and UV, these are merely relative to Saturn's disk.

Table 6

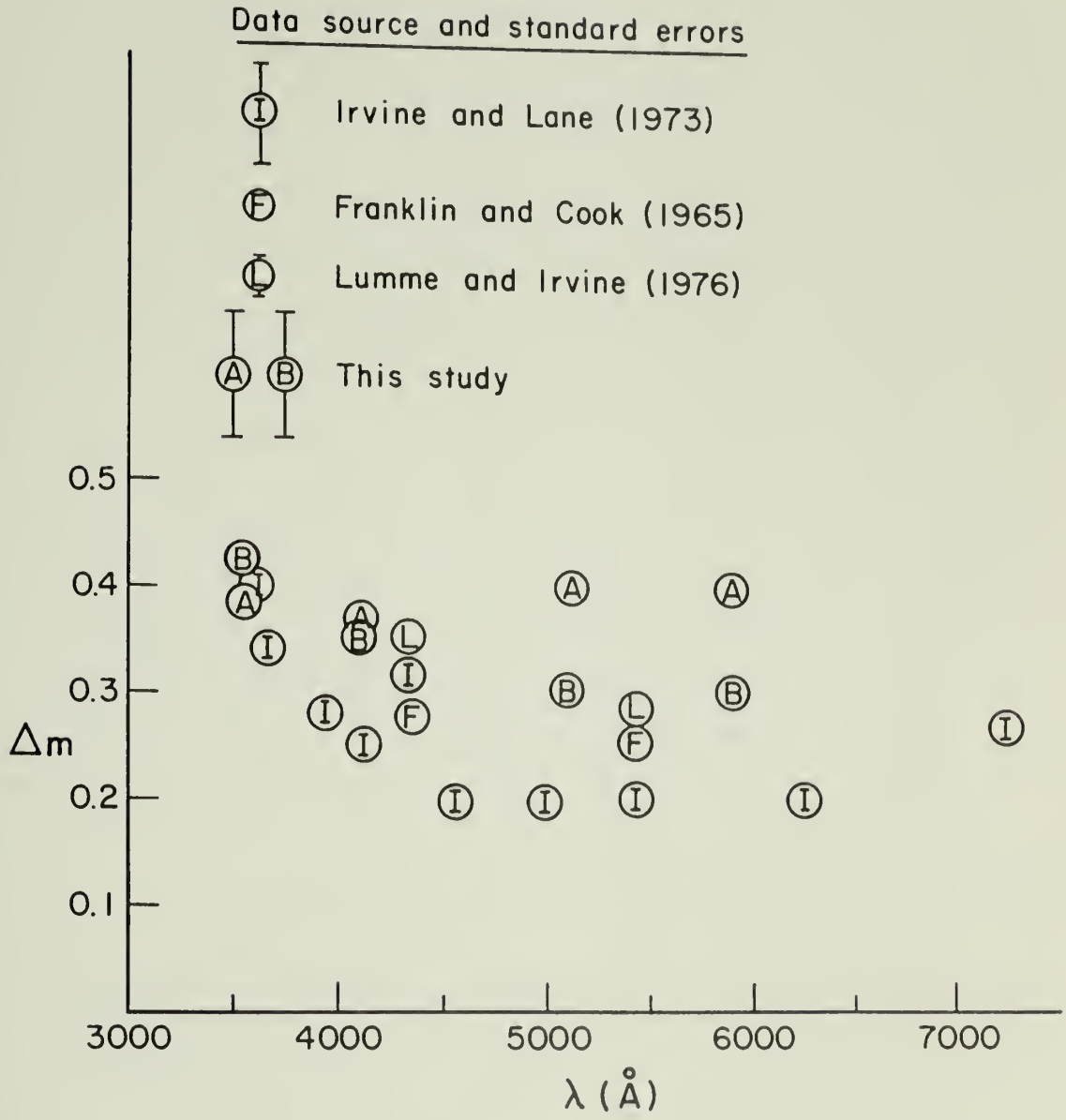
The parameter a_1 from logarithmic fit to all Mauna Kea data, $m(\alpha) = a_0 + a_1 \log \alpha$.

	This study		Lumme and Irvine		Franklin and Cook
Color	Ring A	Ring B	Ring A	Ring B	Both rings together
Red	0.24 ± 0.02	0.19 ± 0.02	0.27 ± 0.04	0.24 ± 0.03	
Green	0.29 ± 0.02	0.23 ± 0.02	0.28 ± 0.04	0.24 ± 0.02	0.251 ± 0.004
Blue	0.24 ± 0.02	0.23 ± 0.02	0.35 ± 0.04	0.34 ± 0.04	0.280 ± 0.003
UV	0.25 ± 0.02	0.25 ± 0.02			

at $B \sim 26°$ (Lumme and Irvine, 1976) we see that this parameter is substantially different only in blue. In green, however, there is good agreement among all observations including Franklin and Cook's photoelectric (V) data. Thus, the shape of the visual phase curve appears to be constant over this range of tilt angle, and consequently well determined. As noted by Lumme and Irvine (1976) the parameter a_1 is consistently larger in ring A than in ring B. This difference is of the same magnitude as in their study and they were able to make a satisfactory explanation for it in terms of the different optical depths of the two rings. The difference is then entirely due to increased multiple scattering, which decreases the phase variation in ring B. Also, note that the difference in this parameter between the two rings decreases as wavelength (and thus albedo) decreases. This is just what would be expected if the explanation by multiple scattering is correct. The difference in the shape of the blue phase curves may indicate a real variation with tilt angle since it requires that the uncertainties be much larger than claimed by any of the authors if the phase curves are actually the same at $\bar{B} \sim 17°$ and $\bar{B} \sim 26°$.

The values for Δm can serve to discriminate between the two rings. This quantity is plotted against previous determinations in Figure 9. To my knowledge, this is the first determination for the magnitude of the opposition effect for ring A alone, although Franklin and Cook's measurements of both rings together are comparable since they could find no significant difference between the rings. In ring A the value of Δm shows no evidence of dependence on color, while for ring B these data show the familiar increase toward shorter wavelengths noted by Irvine

Figure 9. The opposition effect for Saturn's rings.


Data source and standard errors
Irvine and Lane (1973)
Franklin and Cook (1965)
Lumme and Irvine (1976)
This study
0.5
0.4
0.3
0.2
0.1
Δm
3000
4000
5000
6000
7000
λ (Å)

and Lane (1973). In this case the opposition effect is smallest where the ring albedo is greatest, showing the effect of increased multiple scattering in "washing out" the opposition surge. The absolute discrepancy of about 0.1 magnitudes between the present data and that of Irvine and Lane may be explained by their different method for finding Δm and the fact that they measured both rings together photoelectrically. As to differences between the rings, note that Δm for ring A is larger in all colors (except UV) than for ring B, although the standard errors are so large that they overlap at one sigma. This is likewise as expected from the larger amount of multiple scattering in ring B.

From the phase curve data several conclusions can immediately be reached. First, several analyses give a consistent shape for the green phase curve of Saturn's brighter B ring, which is not very different at $\bar{\bar{B}} = 17°$ and $\bar{B} = 26°$. New phase curves in both red and UV have been plotted, but they are relative to the center of the disk and therefore limited by our lack of knowledge of the phase variation of the disk in these colors. Secondly, the differences in the phase curves with color and of the ratio of ring A to ring B are consistent with the different contribution of multiple scattering due to the albedo spectrum of the ring particles. Thirdly, the difference in the behavior of the two rings is consistent with differences mostly in the amount of multiple scattering, in this case due to different optical depths. However, some difference in volume density of scatterers is necessary to explain the similarity of the phase curves at both tilt angles. The next section will justify this claim with detailed multiple scattering models for the green phase curves, for which absolute calibration is possible.

Analysis in Terms of Multiple Scattering Models

Several recent studies of the photometric behavior of Saturn's rings have been made with multiple scattering models. The work of Kawata and Irvine (1975) analyzes the phase curves observed by Franklin and Cook (1965), and incorporates the improvements made since Seeliger's (1887) first explanation of the opposition surge as due to the mutual shadowing of the particles. Their study takes into account 1) the finite size of the sun, 2) the solar limb darkening, 3) possible dispersion in the particles sizes, and 4) the multiple scattering among the particles. In addition, they are able to consistently explain the observations of the infrared brightness (e.g., Murphy, 1974). More recent infrared observations (Nolt, et al., 1977) are also consistent with their model. Visual observations of the brightness variation with changing declination of the earth and sun (the tilt effect) have been analyzed by Lumme (1970), Price (1974), and Esposito and Lumme (1977). For the visual band the analyses of both the phase curves and of the tilt effect are quite consistent. For ring B, they require 1) high single scattering albedo, 2) optical depth of at least unity, 3) backscattering phase functions. In addition, Kawata and Irvine found that the fractional volume occupied by the scatterers in ring B was $0.006 \leq D \leq 0.012$, which is consistent with earlier studies of the phase curves. Esposito and Lumme required that the phase function be moderately forward scattering in addition to backscattering, and found that the particles in ring A need not be different from those in ring B.

Some of the complications of the model of Kawata and Irvine may not

be essential to analysis of the phase curves. For example, they found they could match the phase curves with particles all the same size and that the solar limb darkening produce an insignificant effect. In addition, in the range of D found to match the phase curve, the error in treating the sun as a point is less than 3% (Kawata and Irvine, 1974). In Chapter III, it was shown that the inclusion of shadowing in only the first-order scattering is a very good approximation. Because the present photographic data have considerably more uncertainty than that of Franklin and Cook, it will be sufficient to use a four parameter model for the interpretation, as follows. The rings will be considered a homogeneous layer many particles thick, composed of equal radii spheres as the scattering elements. The sun will be taken as a point source, so that the methods of Chapter III will apply in determining the effect of shadowing. The multiple scattering will be computed by a Markov chain formalism as in Chapter II. The declination of the earth and sun will be taken to be $\bar{B} = 16.74°$ which gives $\mu = \mu_0 = \sin \bar{B} = 0.288$. The parameters of the model are 1) the phase function for scattering $P(\alpha)$, 2) the single scattering albedo $\tilde{\omega}_0$, 3) the total optical depth τ_0, and 4) the fractional volume occupied by scatterers D. The range of parameters which satisfy the present observations of the B ring in green will be compared with the results of previous studies. Then it will be shown that the A ring phase variation can be matched with ring particles having the same scattering properties. Thus, the two rings need only differ in geometry, i.e., in D and τ_0.

The absolute scale of the observations, that is, their conversion to specific intensity, I, relative to the incident solar flux πF, can

be determined by the value of Lumme and Irvine (1976) for the specific intensity of the center of Saturn's disk (equation 47). However, as in Esposito and Lumme (1977), because of the large uncertainty in this determination, the lower bound for disk center brightness will be used, i.e.

$$I_D/F = 0.45.$$

This will allow the largest possible range for the model parameters. If a larger value is taken for I_D/F, that will constrain the albedo to be higher, the phase function to be more backscattering and the allowed ranges for the other parameters to be somewhat smaller, but the results are basically unchanged. For ring A, the brightness will be multiplied by 1.135 to correct for the effect of atmospheric and instrumental smearing which decreases the brightness of the narrow ring A maximum. This is an average figure which has been used by Lumme and Reitsema (1978).

A total of 75 multiple scattering calculations were compared with the observations. The phase function was modeled as a Henyey-Greenstein function (equation 44) where the average of the cosine of the scattering angle had one of the values

$$g = -0.5, -0.4, -0.3, -0.2, 0.0, 0.3, 0.5 \quad .$$

Also tried was the sum of two Henyey-Greenstein functions

$$P(\alpha) = bP(g_1,\alpha) + (1-b)\, P(g_2,\alpha)$$

with

$$\begin{aligned} &1) \quad b = 0.404 \quad g_1 = -0.5 \quad g_2 = +0.5 \\ &\text{or} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (49) \\ &2) \quad b = 0.143 \quad g_1 = -0.7 \quad g_2 = +0.7 \quad . \end{aligned}$$

The 2 sets of parameters in equation (49) had both been found to give a reasonable match to the tilt effect for the B ring by Esposito and Lumme (1977). The albedo for single scattering had the values

$$\tilde{\omega}_0 = 0.75,\ 0.82,\ 0.85,\ 0.90.$$

The total optical depth was taken as

$$\tau_0 = 0.4,\ 0.5,\ 0.8,\ 1.0,\ 1.2,\ 1.5\ .$$

The fractional volume took on variously the values

$$D = 0.001,\ 0.003,\ 0.005,\ 0.006,\ 0.008,\ 0.010,\ 0.0125,\ 0.015,\ 0.020,\ 0.025\ .$$

A particular model calculation was judged to satisfactory fit the data if it lay within 10% of the absolute phase curve at the three points $\alpha = 0.1°$, $1.0°$, $3.6°$, and the ratio of intensity at $\alpha = 0.1°$ and $\alpha = 3.6°$ was in the range

$$\frac{I(\alpha = 0.1°)}{I(\alpha = 3.6°)} = \begin{array}{ll} 1.37 \pm 0.14 & \text{(ring B)} \\ 1.47 \pm 0.15 & \text{(ring A)} \end{array},$$

where the values are from the observations. The latter requirement assures that a calculation was not satisfactory if 10% low at $\alpha = 3.6°$ and 10% high at $\alpha = 0.1°$, for example. These constraints gave fifteen models which matched ring B and four matches to ring A.

For ring B in green light the conclusions are as follows. As might be expected there is little effect due to optical depth as long as $\tau_0 \geq 0.8$. Because the ring is optically thick, we have little discrimination for this parameter. Consistent with almost all previous studies, we find the albedo for single scattering must be quite high, i.e., $\tilde{\omega}_0 \gtrsim 0.80$. Unlike observations of the tilt effect these observa-

tions constrain only the backward scattering portion of the phase function. For various values of the other parameters, suitable fits were found for $g = -0.3$ and the two phase functions (49) from Esposito and Lumme (1977). The smaller the absolute value of g (or g_1), the larger the density D that was required for an optimal fit to the observations. Our results show that the phase functions with $g \leq -0.4$ are too strongly backscattering to match the data and those with $g > -0.2$ are too weakly backscattering. These results are quite consistent with the conclusions of Kawata and Irvine (1975). For the fractional volume occupied by scatterers, possible fits required

$$0.003 < D < 0.020 \quad .$$

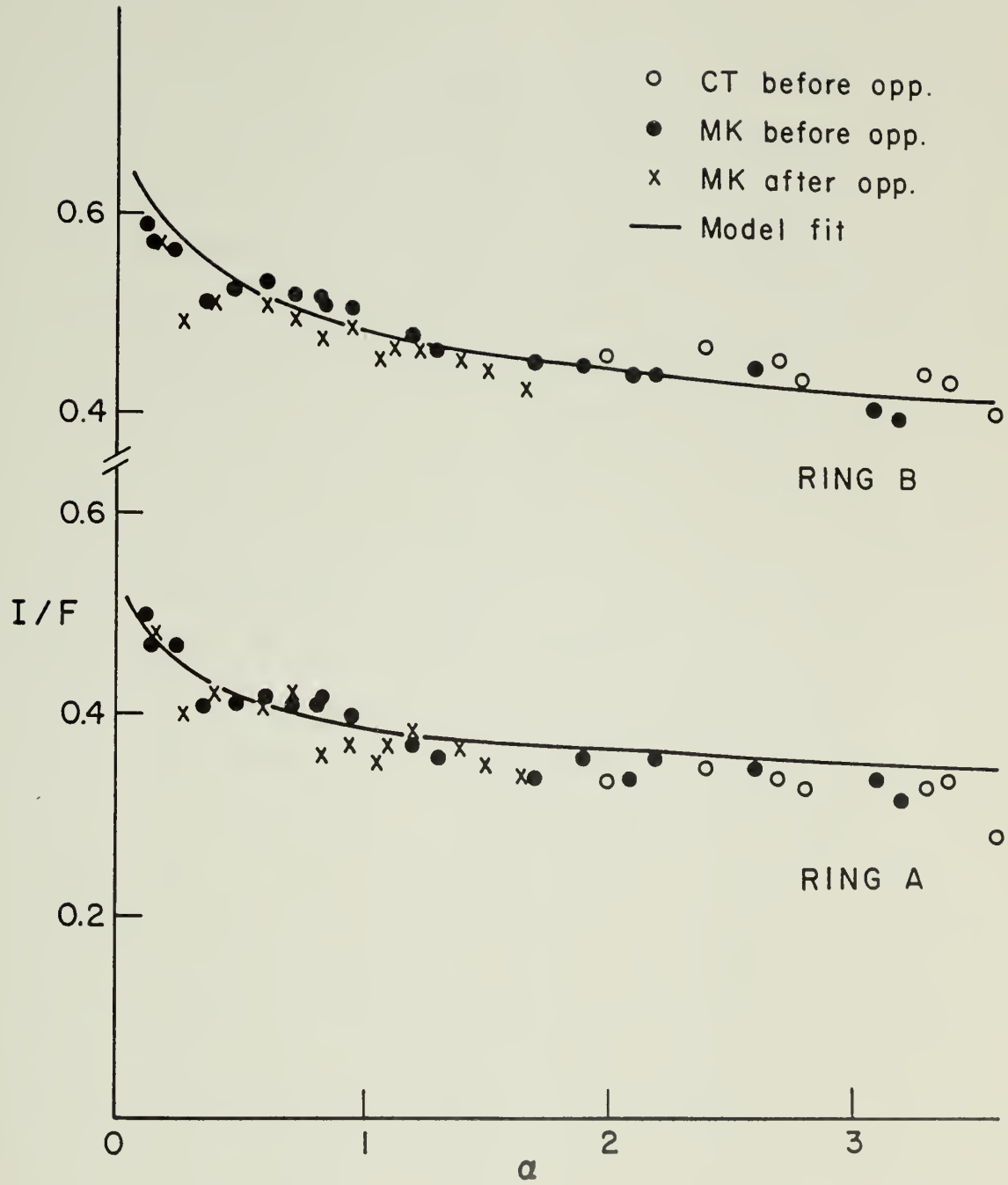
This range brackets the results of Kawata and Irvine though it is somewhat larger due to the larger uncertainty in the present observations.

For ring A and green light, we find fits possible for the same phase functions and albedo, with $\tau_0 < 0.5$ and similar D. For example, a good fit to both rings is found with the model:

$$\begin{aligned} &b = 0.404 \quad g_1 = -0.5 \quad g_2 = -0.5 \quad \tilde{\omega}_0 = 0.9 \ ; \\ &\text{for ring A} \quad \tau_0 = 0.4 \quad D = 0.005 \ ; \\ &\text{for ring B} \quad \tau_0 = 1.0 \quad D = 0.008 \ . \end{aligned} \tag{50}$$

It does not seem that the difference in D is significant. In fact, when fitting each ring without regard to the other, different values of D are found so that D is greater in the A ring. The calculated phase curves from (50) are plotted against the nightly averages of the observations in Figure 10. The fits are not excellent, and it is not surprising that better fits may be found for each ring alone (four free parameters are

Figure 10. Model phase curves for Saturn's rings. For model parameters see equation (50).


CT before opp.
MK before opp.
MK after opp.
Model fit
0.6
0.4
RING B
0.6
I/F
0.4
RING A
0.2
0
1
2
3
α

better than two). However, it does not seem that the quality of the data requires yet that we conclude the scattering particles are different in the two rings. The conclusions for the model parameters are summarized in Table 7. Recall that these parameters apply to the brightest portion of each ring. The limits given for each parameter are the extreme values: a parameter value lies in this range if for any values of the other parameters a suitable fit was possible. Thus, it is not possible to make a satisfactory model which has the extreme values as its parameters. On the other hand, even with four parameters it is not possible to sample all combinations. It may be for a provident selection of the others, the range of one of the parameters might be extended. The limitation on this is that the different parameters are constrained by different parts of the phase curve, as follows. The total optical depth is only slightly constrained by this model, but influences both the size of the opposition effect and the amount of multiple scattering, which affects the slope of the phase curve outside the opposition peak. The fractional volume D has an effect only on the size and shape of the opposition peak. The larger that D is, the broader and higher is the opposition peak. Likewise, the shape of the phase curve outside the peak is determined mostly by the shape of the phase function in the nearly backscattering direction (which gives the single scattering) and the magnitude of the albedo (which decreases the slope due to multiple scattering). As a result, we can be confident that the ranges of the parameters in Table 7 cannot be stretched too far. Recall that these conclusions apply only to the green phase curves. We find that, like the phase curves themselves in green, the conclusions are quite similar at $\bar{B} \cong 17°$ and $\ddot{B} \cong 26°$.

Table 7

Allowed range for the parameters (from green phase curves).

	This study	Esposito and Lumme	Kawata and Irvine
Ring B	$\tau_0 \gtrsim 0.8$	$\tau_0 \gtrsim 0.9$	$\tau_0 > 0.5$
	$-0.4 \leq g < -0.2$	$g_1 \leq -0.3$	[a]
	$\tilde{\omega}_0 \gtrsim 0.8$	$\tilde{\omega}_0 \gtrsim 0.8$	$0.82 \lesssim \tilde{\omega}_0 \lesssim 0.90$
	$0.003 \lesssim D \lesssim 0.020$		$0.006 \lesssim D \lesssim 0.012$
Ring A	$\tau_0 \lesssim 0.5$	$0.4 \lesssim \tau_0 \lesssim 0.6$	
	$0.003 \lesssim D \lesssim 0.015$		
	(other parameters same as ring B)		

[a]Phase functions are similar but quantitative comparison is difficult.

In the other colors, the same methods can be used if we have an absolute scale for the phase curves. This can be found from the observations of Lumme and Reitsema (1977) by the same method as used for the green phase curve. The phase variation of Saturn's disk in red and ultraviolet will be ignored because it is not well known and is not likely to be so large as to make a significant difference over the present range of phase angle. This procedure yields phase curves which can be compared to multiple scattering calculations. These comparisons determine whether the color variation of the phase curves can be explained by differences in particle albedo alone as suggested above. Kawata and Irvine (1975) have found that such an albedo variation was sufficient to explain the blue and visual phase curves for the B ring observed at maximum declination of the ring plane.

For definiteness, we consider the following model of the rings:

$$\begin{aligned} &1)\ \text{single particle phase functions as in (49)} \\ &2)\ \tau_0(A) = 0.4 \\ &\quad\ \tau_0(B) = 1.0 \\ &3)\ D = 0.01\ . \end{aligned} \tag{51}$$

The single scattering albedo is allowed to vary with color in the range $0.3 \leq \tilde{\omega}_0 \leq 0.9$. Model calculations of the phase curves indicate the range of particle albedo that is consistent with observations. The results are not dependent on the choice of phase function and are also quite similar for each ring. In all colors, a satisfactory fit could be found consistent with the data accuracy. The allowed values for $\tilde{\omega}_0$ are in Table 8 along with the results of Kawata and Irvine for the B ring. These detailed calculations serve to bolster the earlier

Table 8

The range of single scattering albedo which is consistent with the color dependence of the phase curves being entirely due to variation in particle albedo. The other parameters for the model are listed in (51). The data of Kawata and Irvine refer only to the B ring.

Color	Ring A	Ring B	Kawata and Irvine
Red	$0.75 \lesssim \tilde{\omega}_0$	$0.80 \lesssim \tilde{\omega}_0$	
Green	$0.75 \lesssim \tilde{\omega}_0$	$0.80 \lesssim \tilde{\omega}_0$	$0.82 \lesssim \tilde{\omega}_0 \lesssim 0.90$
Blue	$0.55 \lesssim \tilde{\omega}_0 \lesssim 0.65$	$0.5 \lesssim \tilde{\omega}_0 \lesssim 0.6$	$0.65 \lesssim \tilde{\omega}_0 \lesssim 0.75$
Ultra-Violet	$0.40 \lesssim \tilde{\omega}_0 \lesssim 0.55$	$0.4 \lesssim \tilde{\omega}_0 \lesssim 0.45$	

analysis of Kawata and Irvine, lend credence to the efficacy of the classical model of the rings as a layer many particles thick, and support the contention that Saturn's rings are formed of particles of quite similar composition.

---

[1] It now appears that this discrepancy may be due to the range of calibration points used in the least squares fit for some images being too small. As a result the true intensity of the rings may be underestimated with respect to the disk on the densest exposures at Mauna Kea, which are in the blue and ultra-violet. The conclusions regarding the shorter wavelength observations, such as the albedos quoted in Table 8, are thus somewhat uncertain, and may be changed by possible later corrections for this effect in connection with related research (Esposito, et al., 1977).

CHAPTER V

CONCLUSION

This thesis has developed a new method to solve radiative transfer problems, based on the theory of Markov chains. This method was tested over a range of parameters and found to be both fast and accurate. Using this formalism for radiative transfer, the classical calculation of the effect of mutual shadowing in diffuse reflection was extended to higher orders of scattering beyond the first. Every photon that escapes the atmosphere has the effect of shadowing included after both its first and last scattering in the atmosphere, and all the higher order of scattering are calculated consistent with the effect of shadowing on the photons scattered once. Other extensions to the classical calculation were to extend its applicability to higher volume density by a "Van der Waals" type approximation, and to allow for some inhomogeneity in the particle distribution which can be described by a pair correlation function.

New phase curves for Saturn's rings were presented from observations at $\bar{B} \sim 17°$, an aspect where good data is lacking. These show some difference in the magnitude of the opposition effect for the two rings. The phase curves in green were analyzed with simple multiple scattering models including the effect of shadowing, using a Markov chain formalism.

The conclusions of this study may be summarized as follows. A Markov chain method for a plane-parallel atmosphere utilizing 200 states gives results accurate to better than 1% in the range $-0.7 \leq g \leq 0.5$,

$\tau_0 \leq 4.0$, $\tilde{\omega}_0 \leq 0.9$. Multiple scattering calculations for Saturn's rings which included shadowing only in the primary scattering are probably accurate to better than 0.1%. The effect of mutual shadowing in the higher order components may not be negligible for a planetary regolith, but further analysis is required before a quantitative study can be made.

Compared to the previous data from $\bar{B} \sim 26°$, the new phase curves are remarkably similar in green. In juxtaposition to the opposition effect for the B ring, ring A shows no color dependence for Δm. The color dependence for the B ring agrees well with the measurements of Δm by Irvine and Lane (1973). The color dependence of the phase curves is as expected for a multiple scattering model with a greater contribution from multiple scattering at the wavelengths where the ring albedo is greatest.

Detailed analysis of the phase curves in green show that results for the B ring are consistent with the recent studies of the tilt effect and the phase curve at $B \sim 26°$. The allowable range of parameters for the scattering model is summarized in Table 7. Although the match is not excellent, it is still possible that the particles have the same scattering properties in both rings. The fact that the same parameters can explain the tilt effect in each ring and give a good match to the phase curves at both $B \sim 17°$ and $B \sim 26°$ in all colors (see Table 8) argues strongly for the classical, many-particle thick model of the rings.

REFERENCES

Alcock, C. and Goldreich, P. (1977). Paper presented at Division for Planetary Sciences Meeting, Honolulu. B.A.A.S. 9, 462.

Baum, W. A. (1973). Planet. Space Sci. 21, 1511.

Bharucha-Reid, A. T. (1960). Elements of the Theory of Markov Processes and Their Applications, McGraw Hill, New York.

Bobrov, M. S. (1940). Astron. Zh. USSR 17, 1-8.

_____. (1956). Astron. Zh. USSR 33, 161; 904.

_____. (1961). Astron. Zh. USSR 38, 669.

_____. (1970). In Surfaces and Interiors of Planets and Satellites (A. Dollfus, ed.), Academic Press, New York.

Brahic, A. (1977). Paper presented at Division for Planetary Sciences Meeting, Honolulu. B.A.A.S. 9, 461.

Chandrasekhar, S. (1950). Radiative Transfer, Oxford University Press, London.

Cheyney, H. and Arking, A. (1976). Ap.J. 207, 808.

Cole, G. H. A. (1959). Advances in Physics 8, 225.

Columbo, G., Goldreich, P., and Harris, A. W. (1977). Paper presented at Division for Planetary Sciences Meeting, Honolulu. B.A.A.S. 9, 462.

Esposito, L. W. and House, L. L. (1978). Ap.J. (in press).

Esposito, L.W., Irvine, W.M., Lumme, K., and Baum, W.A. (1977). In Planetary Atmospheres (A. Vallance-Jones, ed., in press).

Esposito, L. W. and Lumme, K. (1977). Icarus 31, 157.

Franklin, F. A. and Colombo, G. (1977). Center for Astrophysics preprint 685, submitted to Icarus.

Franklin, F. A. and Cook, A. (1965). Astron. J. 70, 704.

Green, C. D. (1969). Integral Equation Methods, Barnes and Noble, New York

Hansen, J. and Travis, L. (1974). Space Sci. Rev. 16, 527.

Hapke, B. W. (1963). J.G.R. 68, 4571.

Hertzsprung, E. (1919). Astr. Nachr. 208, 81.

House, L. L. and Avery, L. (1969). J.Q.S.R.T. 9, 1579.

Irvine, W. M. (1964). Bull. Astron. Inst. Neth. 17, 266.

_____. (1965). Ap.J. 142, 1563.

_____. (1966). J.G.R. 71, 2931.

_____. (1975). Icarus 25, 175.

Irvine, W. M. and Lane, A. P. (1971). Icarus 15, 18.

_____. (1973). Icarus 18, 171.

Kawata, Y. and Irvine, W. M. (1974). In Exploration of the Solar System (A. Woszczyk and C. Iwaniszowska, eds.), Reidel, Dordrecht, Holland.

_____. (1975). Icarus 24, 472.

Kemeny, J. L. and Snell, J. (1960). Finite Markov Chains, Van Nostrand, Princeton.

Lenoble, J. (ed.) (1977). Standard Procedures to Compute Atmospheric Radiative Transfer in a Scattering Atmosphere, Volume I and II. Published by the Radiation Commission of the International Association of Meteorology and Atmospheric Physics at NCAR, Boulder, Colorado.

Lumme, K. (1970). Ap. and Space Sci. 8, 90.

_____. (1971). Ap. and Space Sci. 13, 219.

Lumme, K., Esposito, L. W., Irvine, W. M., and Baum, W. A. (1977). Ap.J. (Letters) 216, L123.

Lumme, K. and Irvine, W. M. (1976). Astron. J. 81, 865.

_____. (1977). Paper presented at Division for Planetary Sciences Meeting, Boston. B.A.A.S. (in press).

Lumme, K. and Reitsema, H. J. (1977). Icarus (in press).

Müller, G. (1893). Publ. Obs. Potsdam 8, 193.

Murphy, R. E. (1974). In The Rings of Saturn (F. Palluconi and G. Pettengill, eds.), N.A.S.A. SP-343.

Nolt, I. G., Tokunaga, A., Gillett, F. C., and Caldwell, J. (1977). Ap.J. (Letters) (in press).

Palluconi, F. and Pettengill, G. (eds.) (1974). The Rings of Saturn, N.A.S.A. SP-343.

Papoulis, A. (1965). Probability, Random Variables and Stochastic Processes, McGraw Hill, New York.

Pollack, J. B. (1975). Space Sci. Rev. 18, 3.

Preisendorfer, R. W. (1965). Radiative Transfer on Discrete Spaces, Pergamon Press, Oxford.

Price, M. J. (1973). Astron. J. 78, 113.

_____. (1974). Icarus 23, 388.

_____. (1977). Icarus 30, 760.

Schoenberg, E. (1922). Ann. Acad. Sci. Fennicae Ser. A, 16.

_____. (1929). In Handbuch der Astrophysik (Herausgegaben von G. Eberhard, A. Kohlschütter, and M. Ludendorf) Band II/1, J. Springer Verlag, Berlin.

_____. (1933). Vjschr. Astr. Ges. Lpz. 68, 387.

Seeliger, H. (1887). Abh. der Bayer. Akad. Wiss. Math.-Naturw. Kl. II 16, 405.

_____. (1895). Abh. der Bayer. Akad. Wiss. Math.-Naturw. Kl. II 18, 1.

Sobolev, V. V. (1975). Light Scattering in Planetary Atmospheres (trans. by W. M. Irvine), Pergamon Press, Oxford.

Takacs, L. (1960). Stochastic Processes, Methuen and Co., Ltd., London.

Varga, R. S. (1962). Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, New Jersey.

Veverka, J., Goguen, J., Yang, S., and Elliot, J. L. (1977). Submitted to Icarus.

Young, D. M. (1971). Iterative Solution of Large Linear Systems, Academic Press, New York.

APPENDIX

MARKOV CHAINS

A summary of some basic concepts from the theory of Markov chains follows. No proofs are given but these can be readily found (see Kemeny and Snell 1960, or Bharucha-Reid 1960, for example). When possible, a physical interpretation will be supplied.

Systems of a probabilistic nature which evolve with time can generally be classed as stochastic processes. Quite generally, a stochastic process is an arbitrary family of real random variables $\{\xi_t, t\varepsilon T\}$. More to the point, J. L. Doob has defined a stochastic process as the mathematical abstraction of an empirical process whose development is governed by probabilistic laws. The term "stochastic process" refers to the mathematical representation and not to the empirical process itself.

A stochastic process may have the property that its subsequent evolution depends only on the present state of the system. In this case, the past history of the system and the exact manner in which it reached its present configuration are irrelevant to predicting its future course. This is called the Markov property, and systems possessing this property are called Markov processes. In terms of conditional probabilities, we say that consecutive trials form a Markov process if for all possible n

$$P(\xi_n = j \mid \xi_0 = i_0, \xi_1 = i_1, \ldots, \xi_{n-1} = i_{n-1}) = P(\xi_n = j \mid \xi_{n-1} = i_{n-1}) .$$

Here, $P(A|B)$ is the probability that A occurs given B. The random variables $\xi_n = j$ if the outcome of the trial n is the event j.

For definiteness, we will assume that the set of events are exhaustive, mutually exclusive, and finite in number. We require that the

conditional probability $P(\xi_{n+1} = j \mid \xi_n = 1) \equiv P_{ij}$ is independent of n. Then the process is a finite Markov chain. With obvious physical implications, we call the possible events states of the system and the above conditional probabilities are called transition probabilities, P_{ij}. The random variable ξ_0 is the initial distribution. If $\xi_{n-1} = i$ and $\xi_n = j$, we say that the system made a transition from state i to state j at the n-th step. The Markov chain is totally specified by its initial distribution and its transition probabilities.

Algebraically, we can represent the initial distribution as a row vector $(\Pi_0)_i = P(\xi_0 = i)$. We can arrange the transition probabilities in a square matrix, P, such that the entry p_{ij} of P is the probability that the system make a transition from state i to state j. This matrix is known as the transition matrix.

From the matrix P, which gives the probability that the system goes from state i to state j in a single step, we can derive the n-step transition probabilities that a system go from state i to state j in n steps. These are simply given by P^n, the n-th power of the matrix P. We find $P(\xi_{m+n} = j \mid \xi_m = i) = (P^n)_{ij}$. Of special interest is the asymptotic behavior of the probabilities as $n \to \infty$. In this regard, it is helpful to classify the states of the chain.

Two states, i and j, are said to communicate if for some n and m there is a non-zero probability of reaching state j from state i in n steps and similarly reaching state i from state j in m steps. This can be shown to be an equivalence relation which thus divides the states into discrete equivalence classes. If all states communicate there is only one class and the chain is said to be irreducible. A class is said

to be closed if it is not possible to reach a state outside the class in a single transition. If such a closed class contains only one state, we say that the state is absorbing.

For a finite Markov chain, a class is either transient or ergodic. Once the system leaves a transient class it never returns. If a system enters a state in an ergodic class, it never leaves that class. In the sense of communication, the transient states are the transmitters and relayers, while the ergodic states are the ultimate receivers. If all the ergodic states of a Markov chain are absorbing, then it is an absorbing chain. It is easily shown that an absorbing chain has a transition matrix which can be put in the following form.

$$P = \begin{bmatrix} I & 0 \\ R & Q \end{bmatrix}$$

if the chain has r absorbing states and n-r transient states,

I is the rxr identity matrix,

0 is the rx(n-r) null matrix,

R is the (n-r)xr matrix for transitions from a transient to an absorbing state, and

Q is the (n-r)x(n-r) matrix for transitions between two transient states.

With such an absorbing Markov chain, for any initial distribution, the system will ultimately reach and remain in an absorbing state. For a given initial distribution, we are interested in the probability of ultimate absorption in each of the absorbing states. Thus, we investigate the behavior of P^n as $n \to \infty$. Successive multiplication yields

$$P^n = \left[\begin{array}{c|c} I & 0 \\ \hline (I + \sum_{k=1}^{n} Q^k)R & Q^n \end{array}\right] \quad \text{. Since } \lim Q^n = 0, \text{ we have}$$

$$\lim_{n\to\infty} P^n = \left[\begin{array}{c|c} I & 0 \\ \hline X & 0 \end{array}\right] \quad \text{where } X = (\sum_{n=0}^{\infty} Q^n)R \quad .$$

But, $\sum_{n=0}^{\infty} Q^n = (I-Q)^{-1}$. Thus, $X = (I-Q)^{-1}R$, or X is the solution to the matrix equation $(I-Q)X = R$ For any initial distribution, Π_0, we have the ultimate absorption probabilities:

$$(\Pi_\infty)_i = (\Pi_0)_i + \sum_{j=r+1}^{n} (\Pi_0)_j \, X_{ji}, \quad \text{for } 1 \leq i \leq r$$

$$(\Pi_\infty)_i = 0 \quad , \quad \text{for } r \leq i \leq n \quad .$$

This solves the problem.