A Photometric Study of the South Polar Terrain of Enceladus

Bereket Daniel Mamo

Minnesota State University, Mankato

Follow this and additional works at: https://cornerstone.lib.mnsu.edu/etds

Part of the Remote Sensing Commons, and the The Sun and the Solar System Commons

Recommended Citation


This Thesis is brought to you for free and open access by the Graduate Theses, Dissertations, and Other Capstone Projects at Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. It has been accepted for inclusion in All Graduate Theses, Dissertations, and Other Capstone Projects by an authorized administrator of Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato.
A Photometric Study of the South Polar Terrain of Enceladus

by
Bereket Daniel Mamo

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
In
Physics

Minnesota State University, Mankato
Mankato, Minnesota
August 2020
August 20, 2020

A Photometric Study of the South Polar Terrain of Enceladus

Bereket Daniel Mamo

This thesis has been examined and approved by the following members of the student’s committee.

________________________________________
Dr. Paul Eskridge, Advisor

________________________________________
Dr. Analía Dall’Asén, Committee Member

________________________________________
Dr. Fei Yuan, Committee Member
Acknowledgements

First and foremost, I would like to thank God for His unwavering love, mercy, and blessings. This work would have not have been completed without the contributions of many people. I would like to sincerely thank my advisor Dr. Paul Eskridge for his guidance, patience, and support throughout the course of this project and my time at Minnesota State University, Mankato (MSU). Moreover, I would like to express my gratitude to him for the freedom and opportunity he afforded me to explore my interests, without which I wouldn’t have been able to find my research niche. Next, I would like to thank my committee members Drs. Analía Dall’Asén and Fei Yuan for critiquing my work and helping me improve it.

I am deeply grateful for the willingness of Lynn Weller for sharing with me her improved camera pointing kernels for Enceladus images and Dr. Frédéric Schmidt for Hapke model fitting software. In addition, I would like to thank Dr. Paul Helfenstein for the helpful correspondence at different stages of the project, Ben Knowles for useful replies to my CISSCAL questions, and the USGS Astrogeology Science Center Discourse forum community for software help.

My heartfelt gratitude goes out to every faculty and staff member of MSU’s Physics and Astronomy Department. I have been one of the many beneficiaries of the safe, friendly, and nurturing environment you have all created in the Department. I also add to this list graduate and undergraduate students who have all been an integral part of the collegial atmosphere in the Department.

I extend my profound appreciation to my family, near and far. I would not be where I am today without your support, encouragement, and sacrifice, so I humbly thank you.
Contents

1 Introduction ............................................................... 1

2 Enceladus ................................................................. 4
   2.1 Introduction ...................................................... 4
   2.2 Geology .......................................................... 6
      2.2.1 South Polar Terrain (SPT) .......................... 7
   2.3 Plume ............................................................ 10
   2.4 Internal structure and heating ............................... 12

3 Hapke Photometric Model ............................................. 15
   3.1 Definitions ...................................................... 16
      3.1.1 Cross section ............................................ 16
      3.1.2 Efficiencies .............................................. 17
      3.1.3 Single-scattering albedo and size-parameter ........ 18
      3.1.4 Particle phase function ............................... 18
      3.1.5 Irregular particles .................................... 19
      3.1.6 Particulate media ..................................... 19
   3.2 Radiative Transfer Equation in sparsely packed media . 24
   3.3 Bidirectional Reflectance ....................................... 27
      3.3.1 Single-scattering ....................................... 28
      3.3.2 Multiple scattering for isotropic scatterers ....... 29
      3.3.3 Anisotropic scattering ................................. 32
   3.4 Effects of increased filling factor ............................ 33
      3.4.1 Loss of Fraunhofer diffraction ..................... 33
      3.4.2 Transmissivity correction ............................ 35
A Photometric Study of the South Polar Terrain of Enceladus
Bereket Daniel Mamo
Master of Science in Physics
Minnesota State University, Mankato
Mankato, Minnesota
2020

Abstract

Saturn’s moon, Enceladus, is a unique body in the Solar System. It is an icy world with ongoing geologic activity that spews out jets of ice particles, water vapor, and other gases from active tectonic fractures (called sulci) near its south pole. Emission from these jets is the dominant source of particles for Saturn’s diffuse E ring, which envelops other icy satellites. The south polar fractures are also sources of tidally generated high heat flow. Accurate determination of this endogenic thermal output requires knowledge of the thermophysical properties of the surface, which are reliably estimated by modeling the photometric behavior across a range of wavelengths. This study attempts to estimate fundamental photometric parameters of three geologic units within Enceladus’ South Polar Terrain: funiscular plains, south polar reticulated plains (SPRP), and Baghdad sulcus flanks. Images acquired in the clear filter ($\lambda = 651 \text{ nm}$) by the Imaging Science Subsystem (ISS) onboard the Cassini spacecraft are used to solve for Hapke photometric model parameters under a Bayesian inversion framework. The angular scattering pattern of particles in geologic units ranged from strongly backscattering for Baghdad sulcus flanks, to moderately backscattering for funiscular plains, to strongly forward scattering for SPRP. The sulcus flank region has the lowest brightness with a single-scattering albedo $\omega = 0.667 \pm 0.001$ while the SPRP has the highest albedo value of one; the funiscular plains have intermediate albedo with $\omega = 0.795 \pm 0.001$ although this value is lower than previously reported albedo values for Enceladus. In addition, the macroscopic roughness parameter revealed the SPRP has the least roughness ($\bar{\theta} = 4.0^\circ \pm 0.1^\circ$) as opposed to Baghdad sulcus ($\bar{\theta} = 21.1^\circ \pm 0.4^\circ$) and funiscular plains ($\bar{\theta} = 37.46^\circ \pm 0.09^\circ$). However, based on particle size distributions,
the sulcus region would be expected have the largest macroscopic roughness. The data set
used lacked observations near zero phase angle, so the opposition effect parameters could
not be constrained. In addition, observations at phase angles above 90° are needed to better
constrain the angular scattering and roughness parameters. The above uncertainty values
are only statistical and do not include systematic errors, which are larger. Nonetheless, the
results show that particles in the different geologic units are photometrically distinct.
1 Introduction

Remote sensing has been an invaluable tool to study planetary surfaces for centuries. One of the earliest uses of remote sensing techniques to study a planetary surface is by Galileo Galilei when he noted that the diffuse scattering of light from the Moon’s surface meant it was not a smooth sphere but rather rough similar to Earth (Galilei 1638). Since then the application of the technique to planetary surfaces has developed to a quantitative science and has revealed much of what we know about Solar System bodies.

The use of photometric observations to study planetary surfaces hinges on the formulation of theories capable of describing the light scattering properties of naturally occurring media. The variation in reflectance with lighting and viewing geometry, specifically incidence angle of incoming irradiance, emission angle of outgoing radiance, and phase angle (the angle between incident and reflected rays of light), is influenced by the physical properties of the scattering surface. Therefore, models that attempt to relate the directional scattering behavior of natural particulate surfaces to the physical characteristics of those surfaces are important. With such models, it would be possible to predict the reflectance of a planetary body at an arbitrary illumination and viewing geometry if its scattering properties are known. However, in most cases the scattering properties are not known; rather they are the missing piece that must be solved for by modeling photometric observations that only partially cover the possible illumination and viewing geometries. Hence, a model that can accurately translate a set of photometric observations at varying geometries into a prediction of the scattering behavior of the reflecting surface is key for understanding the origin and evolution of planetary bodies.

Several photometric models exist that can be used to interpret reflectance data from planetary surfaces. The most widely applied model is that of Hapke (Hapke 1981, 1984, 1986, 2002, 2008, 2012b), which is based on geometric optics and the equations of radia-
tive transfer. It is an analytic, semi-empirical model that includes parameters to account for the single-scattering albedo (ratio of amount of light scattered to the amount scattered and absorbed), surface roughness, porosity, opposition effect (increase in reflectance at small phase angles) and particle phase function, which describes the angular pattern into which light is scattered. The formulation of the Hapke model has been controversial (e.g. Hapke, 1996; Mishchenko, 1994), and laboratory studies to test its ability to correlate model parameters to sample characteristics have had mixed results (e.g. Helfenstein and Shepard, 2011; Shepard and Helfenstein, 2007; Shkuratov et al., 2007; Souchon et al., 2011). Nevertheless, due to its relative simplicity and fast computation, it is widely used to analyze telescopic and spacecraft observations (e.g. Fernando et al., 2013; Verbiscer et al., 2005; Verbiscer and Veverka, 1994) and has become the standard in planetary photometry.

This study applies the Hapke model to analyze data of Saturn’s moon Enceladus taken by Cassini spacecraft’s Imaging Science Subsystem (ISS) instrument. The goal of the study is to estimate the photometric parameters of the South Polar Terrain (SPT) of Enceladus. The SPT is a geologically young region centered near the south pole that is characterized by prominent fractures which are sources of jets of icy particles and gas dominated by water vapor (Porco et al., 2006; Waite et al., 2006). This terrain is also a source of an unusually high thermal emission that is concentrated along the fractures (Spitale and Porco, 2007). Initial estimates of the thermal output from the Cassini Composite Infrared Spectrometer (CIRS) observations put it at 5.8 ± 1.9 GW (Spencer et al., 2006), which is much higher than the expected radiogenically produced power of 0.3 GW (Porco et al., 2006). Subsequent analysis of high resolution CIRS observations covering a broader range of wavelength put the power output at an even higher value of 15.8 ± 3.1 GW (Howett et al., 2011). The difficulty with such measurements though is that the passive emission from solar radiation must be determined and subtracted from the observed emission. Therefore, the true uncertainty of the values may be higher than indicated. In order to achieve an accurate derivation of
the passive emission, the thermal inertia and the bolometric Bond albedo, which are prime
determiners of surface temperatures, must be known. The thermal inertia is a measure of
how well a surface is able to store and emit thermal energy and is found by comparing
the temperature variations due to varying insolation. The bolometric Bond albedo is the
wavelength-integrated fraction of incident light that is scattered into all directions. It is
found by measuring the reflected light over a wide range of wavelengths and phase an-
gles but can also be determined indirectly from surface temperature distributions (Howett
et al., 2010). Howett et al. (2011) used a constant albedo value that was derived indirectly
by comparing passive background spectra, predicted by their thermal model, to observed
spectra. The present work attempts to determine the photometric parameters of the SPT
region with the hope that it will aid in making a direct estimation of the albedo possible.
This could further help in making a more precise measurement of the thermal output as
the thermophysical properties, and hence the passive emission component, can be better
constrained.

This thesis is organized as follows. In section two, some of the important findings
pertaining to the SPT of Enceladus from the Cassini mission are discussed. The chapter
gives a brief overview of the SPT, but does not do justice to the complex processes and
interactions taking place on the satellite. The references therein give a more extensive
review. Section three discusses the theoretical rationale of the Hapke photometric model.
The fourth section describes instrument characteristics and image calibration steps while
section five presents the methodology followed to obtain photometric surface parameters.
The results of the study and the conclusion drawn from them are given in sections six and
seven, respectively.
2 Enceladus

2.1 Introduction

Enceladus (Fig. 1a) is one of the five mid-sized satellites of Saturn orbiting beyond its main rings. It was discovered by William Herschel in 1789 (Herschel, 1790), although little remained known about the satellite until the Voyager spacecraft flybys in the early 1980s. Enceladus’ relatively small diameter of 504 km belies its truly remarkable nature. The Voyager encounters revealed it possesses an extremely high albedo (Buratti and Veverka, 1984) and a diverse set of geologic units, ranging from heavily cratered terrains to smooth, crater-free plains (Smith et al., 1982). Moreover, ground-based observations of the diffuse E-ring showed that its density peaks near the orbit of Enceladus (Baum et al., 1981). These findings, together with the short lifetimes of ring particles inferred from ring particle dynamics studies (Haff et al., 1983; Jurac et al., 2001), strongly suggested Enceladus to be the origin of a replenishing mechanism, perhaps in the form of a geyser-like activity (Haff et al., 1983). However, proof was yet to come.

In the meantime, further analysis of Voyager images showed Enceladus to have a high single-scattering albedo and a more isotropically scattering surface compared to other icy moons (Buratti, 1985). Verbiscer and Veverka (1994) supplemented Voyager images with low phase angle Earth-based observations to determine a low mean slope angle indicating a smooth surface typical of geologically young surfaces of icy satellites. Application of photometric modeling using Hubble Space Telescope (HST) observations also revealed a brighter trailing hemisphere than the leading hemisphere (Verbiscer et al., 2005) consistent with previous telescopic observations (Franz and Millis, 1975; Kouchmy and Lamy, 1975).

Definitive evidence of geologic activity on Enceladus came in July 2005 when multiple instruments onboard the Cassini spacecraft observed active ejection of water vapor and
ice particles from its south pole. Images taken by the Imaging Science Subsystem (ISS) identified a south polar region that is circumscribed by troughs and ridges at 55 °S latitude; this region is distinguished by its albedo and color contrast, elevated temperature, extreme geologic youth, and prominent fractures (dubbed “tiger stripes”) \cite{Porco et al. 2006}. Jets of ice particle plumes were also imaged emanating from this region. Cassini’s Composite infrared Spectrometer (CIRS) detected an anomalous thermal emission coinciding with the tiger stripes \cite{Spencer et al. 2006}. Between 3 and 7 GW of thermal emission was detected from the south polar troughs at temperatures of up to 145 K making Enceladus only the
third solid body (after Earth and Jupiter’s moon Io) in the Solar System whose internal heat can be detected by remote sensing measurements. Other instruments detected water vapor over the south pole (Hansen et al., 2006), which also contained a few percent of CO$_2$, CH$_4$, and CO and/or N$_2$ (Waite et al., 2006), and micrometer-sized dust particles with a concentration consistent with enhanced dust production from the south polar region (Spahn et al., 2006). The combination of data from multiple instruments proved the decades old suspicion of the presence of geologic activity on Enceladus that is responsible for maintaining the E-ring by continuously supply particles.

The following sections give a general overview of some of the observations made by the Cassini mission. These observations have revealed features that make Enceladus a unique body within the Solar System and improved its significance as a potential environment where life could exist.

2.2 Geology

The earliest geological map of Enceladus produced used Voyager’s highest resolution images (≈ 1 km/pixel) covering the Saturn-facing hemisphere and north polar region. It identified four major geologic regions on the basis of crater population and other landforms: cratered terrain, cratered plains, ridged plains, and smooth plains. Some of these regions can further be subdivided based on differences in crater density and tectonic characteristics (Passey, 1983; Plescia and Boyce, 1983; Smith et al., 1982; Squyres et al., 1983). As the Cassini mission started to provide more complete, high resolution images, additional geological maps were produced. An initial mapping by Spencer et al. (2009) using image data available at the time (Roatsch et al., 2008) identified four geologic provinces: cratered terrain, western (leading) hemisphere fractured plains, eastern (trailing) hemisphere fractured plains, and the South Polar Terrain (SPT). The imaging data used had high resolution cover-
age for the southern and equatorial trailing hemisphere, so the mapping was more focused on the SPT and did not consider subunits for the other regions. Crow-Willard and Pappalardo (2015) rectified this using better resolution images, and their map also divided the surface into four geologic provinces: cratered terrain, trailing hemisphere terrain, leading hemisphere terrain and the SPT. The boundaries of these provinces are broadly consistent with those defined by (Spencer et al., 2009) but vary in detail. Each of the provinces was further divided into subunits based on morphological and structural features. The present study focuses on the south polar region, so only SPT is discussed below.

2.2.1 South Polar Terrain (SPT)

The SPT province is a geologically active region centered near the south pole that contains a varied set of terrain features. Crow-Willard and Pappalardo (2015) divide this province into two general units (southern curvilinear unit and central south polar unit; Fig. 2), although Spencer et al. (2009) give a more detailed differentiation (Fig. 3). The major features within the SPT are described below following the (Crow-Willard and Pappalardo, 2015) classification scheme.

- Southern curvilinear unit: This unit is an annular region of ridges and troughs surrounding the central south polar unit. It is bounded by a circumpolar system of south-facing scarps at about 55 °S, which stands ~1 km higher than the interior of the SPT (Patterson et al., 2018). The annulus of ridges and troughs generally trend subparallel to the unit’s outer boundary, which undulates northward and southward. At longitudes where the SPT borders cratered plains to the north, scarps intersect to form dihedral cusps (Spencer et al., 2009), and in the trailing hemisphere these develop into Y-shaped discontinuities that interrupt the circumpolar scarp (Porco et al., 2006). The Y-shaped discontinuities are defined by a pair of scarps that bend north-
ward as the gap between them narrows until they become crudely parallel to each other (Fig. 4a). They transition to a series of N-S trending rifts and fractures that extend towards the equator. The scarps confine a wedge-shaped region consisting of curved, sub-parallel ridges and troughs that are convex-northward (Helfenstein, 2015; Spencer et al., 2009).
• Central south polar unit: This unit is characterized by the tiger stripes, which are linear depressions about 500 m deep, 2 km wide, and 130 km long (Fig. 4b). They are flanked on both sides by 100 m-high ridges. The four major stripes (Alexandria, Cairo, Baghdad, and Damascus Sulci) are arranged *en echelon* and are offset by about 45° from the Saturn direction (longitude 0°). In the Saturn-facing hemisphere, they branch into dendritic patterns while in the anti-Saturn hemisphere (longitude 180°) they form hook-like bends (Porco et al. 2006).

The region in-between the tiger stripes is defined by curved, knobby ridges and elongated, barrel-shaped knobs classified as funiscular (i.e. ropy) ridges by Spencer et al. (2009). These ridges are typically oriented parallel to the adjacent tiger stripes, although they deviate to form quasi-circular patterns in some locations. Broadband ISS spectra and Cassini Visual and Infrared Mapping Spectrometer (VIMS) observations show that the funiscular plains are covered with fine-grained ice particles while the
In addition to the above terrain types, Spencer et al. (2009) define a South Polar Reticulated Plains region that surrounds the tiger stripes and the funiscular plains. It is characterized by cross-cutting fracture patterns at varying orientations.

### 2.3 Plume

One of the most defining feature of Enceladus is the plume emission feeding the E-ring. Cassini images have revealed multiple distinct, prominent jets emanating from the south.
The source locations of the most prominent jets, inferred from triangulation, are the four tiger stripes with Baghdad and Damascus being the strongest sources (Spitale and Porco, 2007). The plume is composed of gas particles, solids (dust), and ions. The neutral gas is the most abundant component, with an average mass flux of 170-250 kg/s (Hansen et al., 2017). Estimates for the flux in solid grains varies but is thought to be up to ~50 kg/s (Ingersoll and Ewald, 2011). The ejection speed of the plume can be estimated from its vertical distribution as well as the degree of collimation of the jets. The dust grains have a wide range of speeds but only about 10% escape Enceladus into the E-ring (Ingersoll and Ewald, 2011). The gas plume has a much larger speed than that of the solid component (Schmidt et al., 2008) and so almost completely escapes into space. The ionic component also escapes Enceladus’ gravitational influence as it gets picked up by Saturn’s magnetosphere.

All three phases of the plume are predominantly composed water. The neutral gas component has water abundance of about 96%. Other species that are found this phase are CO$_2$, NH$_3$, CH$_4$, and H$_2$ with mixing ratios ranging from a fraction of a percent to slightly over a percent (Waite et al., 2017). Smaller quantities of heavier hydrocarbons and organic molecules have also been detected (Waite Jr et al., 2009). Dust grains with radii greater than 0.2 μm have water abundance reaching 99%. Analysis of particle properties from the tiger stripe eruptions indicate that most of the ice grains are in crystalline state (Dhingra et al., 2017). The main non-icy constituents in ice grains (radii >0.2 μm) are sodium salts and organic materials, with potassium salts also present. Salt-rich ice particles dominate the total mass flux of ejected solids but are depleted in the population escaping into the E-ring (Postberg et al., 2011). The proportion of water in the ionic phase is less constrained. However, ions measured in the plume are almost exclusively water and water products (Tokar et al., 2009).

The plume is known to be time variable. Individual jets turn on and off, but this does not
appears to be related to the orbital phase of the Enceladus. However, the plume brightness from emitted ice grains does show a correlation with orbital position of the moon. The brightness peaks when Enceladus is near apocenter (farthest from Saturn) (Hedman et al., 2013), a result that is consistent with geophysical models which predict the tiger stripes open and close in response to temporal variations in tidal stresses (Hurford et al., 2007).

2.4 Internal structure and heating

Knowledge on the internal structure of a planetary body puts constraints on its origin and evolution. This is of particular interest for Enceladus because such knowledge can explain the mechanism of the ongoing geologic activity. Models for the interior structure of Enceladus mainly come from shape, topography, and gravity measurements. The bulk density of Enceladus is 1609 kg m$^{-3}$ (Porco et al., 2006). Based on this value, and its icy surface, Enceladus is thought to consist of a silicate core that is overlain by an ice/water layer (Schubert et al., 2007). Its synchronous rotation—where the satellite shows only one face to Saturn due to tidal locking—means its shape is distorted as it is susceptible to elongation along the tidal axis and centrifugal flattening. Therefore, it takes on a triaxial ellipsoid figure. However, its shape is not one of a body in hydrostatic equilibrium. For ellipsoidal axis lengths $a$, $b$, and $c$, the characteristic ratio $(a-c)/(b-c)$ of a synchronous satellite in hydrostatic equilibrium is 4 (Hemingway et al., 2018). However, based on shape measurements (Thomas, 2010), this value for Enceladus is 2.8, which is clearly non-hydrostatic. Gravity field measurements (Iess et al., 2014), however, show only a modest departure from hydrostatic nature. These results suggest a compensation of the topography. Without compensation, a large non-hydrostatic topography would give rise to a correspondingly large non-hydrostatic gravity field. The most straightforward explanation of the observations is the presence of a low viscosity, high density fluid beneath the crust such as a subsurface
liquid ocean (Hemingway et al., 2018).

Indeed, several other observations support this assertion. The detection of sodium and potassium salts in the ejected ice grains from jets are most naturally explained by the presence of liquid water that has interacted with a silicate core. If the particles had formed by condensation of vapor, which had sublimed from ice, they would have been salt-free. The ratio of solids to vapor detected in the jets also supports the existence of a subsurface liquid. This value has been determined to be $\sim$0.4-0.7, which is a high value that cannot be explained by the adiabatic expansion of vapor alone. The high value could be explained if flash freezing of liquid droplets also played a role (Johnson et al., 2014; Spencer et al., 2018).

Because of its high albedo, the surface temperature of Enceladus is about 80 K at the equator and lower toward the poles. Measurements made by Cassini’s CIRS instrument have shown that the SPT is a region of locally high heat flow that is concentrated along the tiger stripes. The first estimate of the endogenic thermal emission made from observations taken in the wavelength range $\sim$9-16 μm gave a value of $5.8 \pm 1.9$ GW (Spencer et al., 2006). Howett et al. (2011) used CIRS observations made at $\sim$16-1000 μm and determined an excess heat flow that is $\sim$3 times higher, $15.8 \pm 3.1$ GW. Despite the discrepancy of the thermal emission derived from these studies, both are much larger than the $\sim$0.3 GW of radiogenic heat that would be expected from Enceladus given its size, density, heating rate, and fraction of rock makeup.

An obvious question that arises, then, is what could be the source of such a high heat flow. Given the absence of other plausible mechanisms, the dominant source of this excess heat is thought to be tidal heating. Enceladus is in a 2:1 mean-motion resonance with another of Saturn’s mid-sized moon Dione. This means Enceladus orbits Saturn twice each time Dione orbits once. The gravitational perturbations from this configuration forces Enceladus to orbit at an eccentricity value of 0.0047. A slightly eccentric orbit means it
experiences a tidal force that varies with its distance to Saturn during its 1.37-day orbit. Therefore, the amplitude of the tidal bulge experienced by Enceladus also varies. It is the internal frictional resistance to the changing amplitude that results in tidal heat. However, the details of the tidal dissipation mechanism are not fully understood.
3 Hapke Photometric Model

Most of what we know about the surfaces of Solar System bodies comes from remote sensing measurements. Photometric and spectroscopic measurements are widely used in planetary remote sensing to infer information on surface characteristics such as compositional and structural makeup. This is accomplished by the use of models that describe the scattering of light as it interacts with particulate media, like planetary regolith or laboratory powders. There are two methods employed to solve this problem. The first method starts from Maxwell’s electromagnetic equations and attempts to find exact numerical solutions. Although it is possible to find exact solutions for simplified media, modern computational capabilities are not advanced enough to solve Maxwell’s equations for light propagating in complex media like planetary regolith. The second approach is based on the radiative transfer equation (RTE), which describes the balance of radiative energy propagating through a media by accounting for the gain and loss caused by absorption, emission, and scattering (Howell et al., 2015). This approach presents its own challenges as the RTE is itself intractable. For this reason, numerical methods, such as the Monte Carlo method and the doubling method (Van de Hulst, 1980), must be used when a high degree of accuracy is desired. However, due to the computational burden associated with such models, approximate analytic solutions are preferred for photometric analysis by many in the remote sensing community.

Several approximate models have been proposed for the quantitative interpretation of spectrophotometric data of planetary surfaces (e.g. Hapke, 1981, 1984, 1986, 2002, 2008; Lumme and Bowell, 1981). The most popular and developed model is that proposed by Hapke. The Hapke model is based on the RTE and geometric optics. The most complete version of the model has parameters that account for surface roughness, porosity, grain scattering properties, and the opposition effect (the phenomenon of the surge in brightness
that occurs at small phase angles). Comparison with exact solutions of the RTE reveal
that the Hapke model can be accurate to within 10% (Cheng and Domingue, 2000) and, in
general, has high fidelity in fits to photometric data. However, tests to determine the cor-
relation of model parameters to physical characteristics of planetary surfaces have yielded
mixed results (Shkuratov et al., 2007). While there is evidence for a qualitative relation
between some model parameters and regolith properties, a quantitative correlation has not
been firmly established. Furthermore, the ability to find unique solutions in the presence
of several parameters is an added challenge. Despite such shortcomings, the Hapke model
has become the standard for analyzing photometric data of planetary surfaces.

In the following sections, the approximate analytic solution to the RTE will be presented
following the derivation in the original works by Hapke (1981, 1984, 1986, 2002, 2008,
2012b). In section 3.1 definitions of commonly used quantities in single particle scattering
will be given. These are then extended to problems of scattering by particulate media.
In section 3.2 the RTE will be introduced in the context of scattering in a horizontally
stratified medium. Then in section 3.3 an expression for the bidirectional reflectance,
declared as the ratio of the scattered radiance at a detector to the incident irradiance from a
collimated source, will be given for a medium of dispersed particles. The expression will
later be modified to account for the changes that occur when particles are brought close
together. Other simplifying assumptions will also be made, which will later be relaxed.

3.1 Definitions

3.1.1 Cross section

Let the incident light impinging on a particle be \( \mathbf{J} = J \hat{u}_p \), where \( J \) is the irradiance (power
per unit area) and \( \hat{u}_p \) is a unit vector in the direction of propagation. Let the amount of
power that is affected by the particle be $P_E$. Then the **extinction cross section** is

$$\sigma_E = \frac{P_E}{J} \quad (1)$$

A portion $P_S$ of $P_E$ is scattered in all directions, and the remaining power $P_A$ is absorbed by the particle. Thus the **scattering cross section** is given by

$$\sigma_S = \frac{P_S}{J} \quad (2)$$

and the **absorption cross section** is

$$\sigma_A = \frac{P_A}{J} \quad (3)$$

These cross sections are in area units. Since $P_E = P_S + P_A$, it holds that $\sigma_E = \sigma_S + \sigma_A$.

### 3.1.2 Efficiencies

The **extinction, scattering and absorption efficiencies**—$Q_E$, $Q_S$, and $Q_A$—of a particle are given by the ratio of its corresponding cross section to its geometrical cross section. If the particle has a general cross sectional area given by $\sigma = \pi a^2$, where $a$ is the particle radius, then the different efficiencies are given by

$$Q_E = \frac{\sigma_E}{\sigma}$$

$$Q_S = \frac{\sigma_S}{\sigma}$$

$$Q_A = \frac{\sigma_A}{\sigma} \quad (4)$$

It follows from above that the efficiencies must satisfy the relation $Q_E = Q_S + Q_A$. 
3.1.3 Single-scattering albedo and size-parameter

The single-scattering albedo of a particle, $\bar{\omega}$, is the fraction of the extinguished light that is due to scattering:

$$\bar{\omega} = \frac{P_S}{P_E} = \frac{Q_S}{Q_E} = \frac{\sigma_S}{\sigma_E}$$  \hspace{1cm} (5)

The scattering behavior of particles depends on their size relative to the wavelength of the illuminating light, so it is beneficial to define a size parameter $X$ as the ratio of the circumference of the particle to the wavelength

$$X = \frac{2\pi a}{\lambda} = \frac{\pi D}{\lambda}$$  \hspace{1cm} (6)

where $D = 2a$ is the particle diameter.

3.1.4 Particle phase function

The particle phase function describes the angular scattering pattern of $P_S = J \sigma Q_S$. If the power scattered by a particle from direction $\Omega_o$ to direction $\Omega$ is $\frac{dP_S}{d\Omega} (\Omega_o, \Omega)$, then the particle phase function is given by

$$\frac{dP_S}{d\Omega} (\Omega_o, \Omega) = J(\Omega_o) \sigma Q_S \frac{\Pi(\theta)}{4\pi}$$

where $\theta$ is the scattering angle, i.e. angle between $\Omega_o$ and $\Omega$. Since the integration of $\frac{dP_S}{d\Omega} (\Omega_o, \Omega)$ across all directions must give the total scattered power, $P_S$, the particle phase function must satisfy the following normalization condition

$$\frac{1}{4\pi} \int_{4\pi} \Pi(\theta) d\Omega = 1$$  \hspace{1cm} (7)

where $d\Omega = \sin(\theta)d\theta d\psi$. In the case of azimuthal symmetry, the normalization condition
becomes \( (1/2) \int \Pi(\theta) \sin(\theta) d\theta \). The particle phase function for an isotropic scatterer is 1.

### 3.1.5 Irregular particles

The meanings of the above definitions are clear for spherically symmetric particles. However, quantities such as particle radius \( a \) or the extinguished power \( P_E \) do not have a clear meaning for irregularly shaped particles and will likely have different values depending on the orientation of the particle. To extend the definitions for irregular particles, the quantities can be taken to be averages at random particle orientations. For instance \( P_E \) would be the average power extinguished by a particle as it is randomly oriented in different directions. Similarly, the geometric cross section, \( \sigma \), is the average area of the shadow of a particle when oriented at random in all directions. The corresponding particle radius is then \( a = \sqrt{\sigma/\pi} \).

### 3.1.6 Particulate media

The definitions introduced thus far are applicable to single particles. However, in most applications of reflectance spectroscopy, radiative transport in a medium made up of an infinitely thick layer of large particles is of greater interest. Therefore, it is important to extend the above definitions to a collection of particles. Consider a slab with area \( dA \) and thickness \( ds \) (Fig. [5]), containing identical particles that are much larger than the wavelength, separated by distances that are random but are on average larger than the particle size. The volume of the slab is much larger than the volume of a single particle, but the thickness \( ds \) is small enough that no shadowing occurs between particles. The particles are randomly oriented and positioned. Since the separation distance between particles is large, their properties may be considered as though they were isolated.

Let \( N \) be the number of particles per unit volume. Then the total number of particles in the slab is \( N ds dA \), and the power extinguished by these particles is \( dP_E = dI(s) dA d\Omega = \)
Figure 5: Extinction by sparsely distributed particles from Hapke (2012b)

\[-I(s)NdsdA\Omega\sigma Q_E,\] where \(I(s)\) is the radiance. For a sufficiently small \(dI\) and \(ds\),

\[
\frac{dI(s)}{ds} = -I(s)N\sigma Q_E
\]

which may be integrated to

\[
I(s) = I(0) \exp \left( -\int_0^s N(s')\sigma Q_E ds' \right)
\] (8)

The extinction coefficient, \(E\), for a particulate medium can then be defined as

\[
E = N\sigma Q_E
\] (9)

More generally, suppose the slab contains a mixture of particles that could differ in size, shape, composition, and structure. Each type of particle \(j\) will have a number density \(N_j\), geometric cross section averaged over different orientations \(\sigma_j\), volume \(v_j\), and extinction, scattering, and absorption efficiencies \(Q_{Ej}, Q_{Sj},\) and \(Q_{Aj}\), respectively. The total number
of particles per unit volume is then

\[ N = \sum_j N_j \]  \hspace{1cm} (10)

and the average distance between particles is

\[ L = N^{-1/3} \]  \hspace{1cm} (11)

The fraction of space within the medium occupied by particles is the filling factor, \( \phi \)

\[ \phi = \sum_j N_j \nu_j = N \nu \]  \hspace{1cm} (12)

where \( \nu = \phi / N \) is the average volume of a particle. The porosity function is given by \( 1 - \phi \).

The volume-averaged particle cross sectional area is

\[ \sigma = \frac{\sum_j N_j \sigma_j}{N} \]  \hspace{1cm} (13)

The volume extinction coefficient of a particulate medium can then be written as

\[ E = \sum_j N_j \sigma_j Q_{Ej} = N \sigma Q_E \]  \hspace{1cm} (14a)

where

\[ Q_E = E / N \sigma \]  \hspace{1cm} (14b)

is the volume-averaged extinction efficiency. The volume averaged scattering and absorption coefficients and efficiencies can be defined in a similar way as

\[ S = \sum_j N_j \sigma_j Q_{Sj} = N \sigma Q_S \]  \hspace{1cm} (15a)
where

\[ Q_s = S/N\sigma \]  

(15b)

\[ A = \sum_j N_j \sigma_j Q_{Aj} = N\sigma Q_A \]  

(16a)

where

\[ Q_A = A/N\sigma \]  

(16b)

Since \( Q_{Ej} = Q_{Sj} + Q_{Aj} \), it holds that \( Q_E = Q_S + Q_A \) and \( E = S + A \).

The volume angular scattering coefficient, \( G(s, \Omega', \Omega) \) or \( G(g) \), describes the probability that a photon traveling in a direction \( \Omega' \) is scattered into a direction \( \Omega \), and is given as

\[ G(g) = \sum_j N_j \sigma_j Q_{Sj} \Pi_j(g) = N\sigma Q_S p(g) = S p(g) \]  

(17a)

where \( \Pi_j(g) \) is the phase function of particle type \( j \), \( g \) is the angle between \( \Omega' \) and \( \Omega \), and

\[ p(g) = G(g)/S \]  

(17b)

is the volume-average single-particle phase function, i.e. analogous to \( \Pi(g) \) for a single particle. Since \( p(g) \) must satisfy the normalization condition given by eq. (7),

\[ \frac{1}{4\pi} \int_{4\pi} p(g) d\Omega' = 1 \]  

(17c)

and it follows that

\[ S = \frac{1}{4\pi} \int_{4\pi} G(g) d\Omega' \]  

(17d)

For isotropic scatterers, \( p(g) = 1 \).
The *volume single scattering albedo* is

\[ w = S/E \]  

(18a)

and a related quantity called the *albedo factor* is defined as

\[ \gamma = \sqrt{1 - w} \]  

(18b)

The *extinction mean free path* \( \Lambda_E \) is the average distance a photon travels before being extinguished due to absorption or scattering and is given by

\[
\Lambda_E = \frac{\int_0^\infty s' \exp \left[ - \int_{s'}^\infty E_s' ds' \right]}{\int_0^\infty \exp \left[ - \int_{s'}^\infty E_s' ds' \right]} = \frac{1}{E}
\]

(19)

For a continuous medium, the summations in the above definitions are simply replaced with integrals. Thus, with these definitions, a discontinuous medium can be considered to be quasi-continuous. However, because absorbers and scatterers are localized and not smoothly distributed, the physical interpretation of the coefficients is slightly changed. In a continuous medium, the wavefront is affected equally, and thus the intensity is attenuated the same way at any scale. However in a discontinuous medium, the parts of the wavefront that meets the scatterers and absorbers are attenuated while those that pass in between are unaffected. Thus, in such cases it is the intensity over scales much larger than the wavelength and particle separation that is considered to be exponentially attenuated. If the medium consists of different types of particles, the above quantities should be averaged over distances that carry a representative sample of all particles in the medium.
3.2 Radiative Transfer Equation in sparsely packed media

The problem to be solved is the radiance of light received by a detector viewing a horizontally stratified, optically thick medium. Let the $z$ axis be perpendicular to a plane surface separating two spaces at some finite value $z_o$ (Fig. 6). The $z > z_o$ is empty except for a distant source of collimated irradiance $J$ illuminating a large area of the surface and a detector viewing a small area $\Delta A$ of the illuminated surface. The $z < z_o$ space contains irregularly shaped, randomly oriented particles that are large compared to the wavelength. The particle distribution is arbitrary, but their separation distance is assumed to be large enough that no shadowing occurs amongst particles. Since the main interest is planetary regolith, the medium in which particles are immersed in will be assumed to have an index of refraction of unity.

The incoming irradiance makes an angle $i$ with the positive $z$-axis and is at azimuth $\psi = 0$. The detector’s sensitive area, $\Delta a$, is located at a distance $R_o$ (not shown in Fig. 6) from $\Delta A$. It responds to light within solid angle $\Delta \omega$ that makes an angle $e$ with the vertical. The angle between the direction to the source and the detector at the scattering element is the phase angle $g$.

Let $I(s, \Omega)$ be the radiance. It is defined as the power per unit wavelength at point $s$ through a unit area perpendicular to the direction of travel per unit solid angle around the direction $\Omega$. Although $I(s, \Omega)$ is a function of wavelength, the dependence will not be shown explicitly for conciseness. The RTE that describes the propagation of light through the medium is then

$$\frac{\partial I(s, \Omega)}{\partial s} = -E(s, \Omega)I(s, \Omega) + \frac{S(s, \Omega)}{4\pi} \int_{4\pi} I(s, \Omega')p(s, \Omega', \Omega)d\Omega' + \frac{J}{4\pi}S(s, \Omega)p(s, \Omega_i, \Omega) \exp \left[- \int_{s_o}^{s} E(s', \Omega_i)ds' \right] + F_T(s, \Omega)$$

(20)
Figure 6: Schematics of scattering of light from a semi-infinite medium from Hapke (2012b)

where $ds$ is the increment of path length along $\Omega$ toward the detector and $s_o$ is the point along the path length separating the empty half space from the medium. The first term on the right-hand side is the extinction term, describing the decrease in power due to scattering and absorption as the radiance propagates through the scattering element. However, scattering can also increase the power into a beam, since light of intensity $I(s, \Omega')$ propagating through a volume element in a direction $\Omega'$ can be scattered into direction $\Omega$; the second term is this addition due to multiple scattering. The last two terms are the emission terms due to singly scattered incident irradiance from a source and thermal emission, respectively. Specifically, the third term describes light of irradiance $J$ coming from the source that is attenuated by a factor $\exp\left[-\int_{s_o}^{s} E(s', \Omega_i) ds' \right]$ as it is transmitted through the medium and is scattered once towards the detector by the scattering element; $F_T(s, \Omega)$ is the thermal emission term.
The path length can be written in terms of the vertical distance as \( ds = dz / \cos(e) = dz / \mu \), where \( \mu = \cos(e) \). Making this substitution and dividing eq. (20) by \( E(z, \Omega) \), the RTE equation can be written as

\[
\frac{\mu}{E(z, \Omega)} \frac{\partial I(z, \Omega)}{dz} = -I(z, \Omega) + \frac{w(z, \Omega)}{4\pi} \int_{4\pi} I(z, \Omega') p(z, \Omega', \Omega) d\Omega' + \frac{J}{4\pi} w(z, \Omega) p(z, \Omega_i, \Omega) \exp \left[ -\frac{1}{\mu_o} \int_{z}^{z_o} E(z', \Omega_i) dz' \right] + \frac{F_T(z, \Omega)}{E(z, \Omega)}
\]

where \( \mu_o = \cos(i) \). In most applications, including for ensembles of randomly oriented particles, the coefficients \( E, S, \) and \( A \) are independent of \( \Omega \), and the volume-averaged single-particle phase function \( p \) is dependent only on the angle between \( \Omega' \) and \( \Omega \), not on the specific directions of the incoming and outgoing rays. Thus, these quantities can be written as \( E(z), S(z), A(z), \) and \( p(z, g) \), where \( g \) is the angle between \( \Omega' \) and \( \Omega \).

Define the **optical depth** \( \tau \), as

\[
\tau = \int_{z}^{z_o} E(z') dz'
\] (21a)

so that

\[
d\tau = -E(z) dz
\] (21b)

The optical depth is a dimensionless quantity and can be thought of as the number of mean free paths that exist from altitude \( z \) to \( z_o \). Note that according to this definition the optical depth is zero in the empty half space. Combining the above expression for \( \tau \) with eq. (8), light originating from altitude \( z \) is attenuated by a factor \( e^{-\tau} \) when it reaches the surface at \( z_o \).
The RTE can therefore be written in terms of the optical depth as

\[-\mu \frac{\partial I(\tau, \Omega)}{\partial \tau} = -I(\tau, \Omega) + \frac{w(\tau)}{4\pi} \int_{4\pi} I(\tau, \Omega')p(\tau, g')d\Omega' + \frac{J}{4\pi} w(\tau)p(\tau, g) \exp\left[-\frac{\tau}{\mu_o}\right] + \frac{F_T(\tau, \Omega)}{E(\tau)}\]

When written in this form, the RTE depends on \(E, S,\) and \(G\) through the ratios \(w\) and \(p\). In most cases, these parameters are dependent on the altitude through a common function of \(z\). Hence, the ratios \(w\) and \(p\) can be taken to be independent of \(\tau\). A full derivation of the RTE can be found in Hapke (2012b).

### 3.3 Bidirectional Reflectance

Consider the volume element \(dV\) within solid angle \(\Delta\omega\) which is at a distance \(R\) from the detector (Fig. 6). This element is illuminated by scattered light \(I(z, \Omega')\) coming from direction \(\Omega'\) in the medium and by collimated light \(J\) from the source that is transmitted through the medium. The portion of the incoming light that is scattered towards the detector in the direction \(\Omega\) is given by

\[dI = -\left[\frac{w(\tau)}{4\pi} \int_{4\pi} I(\tau, \Omega')p(\tau, g')d\Omega' + \frac{J}{4\pi} w(\tau)p(\tau, g) \exp\left[-\frac{\tau}{\mu_o}\right] + \frac{F_T(\tau, \Omega)}{E(\tau)}\right] \frac{d\tau}{\mu}\]

This light is extinguished by a factor \(\exp\left(-\frac{\tau}{\mu}\right)\) as it traverses the medium before getting to the detector, so the radiance reaching the detector is

\[dI_D = -\left[\frac{w(\tau)}{4\pi} \int_{4\pi} I(\tau, \Omega')p(\tau, g')d\Omega' + \frac{J}{4\pi} w(\tau)p(\tau, g) \exp\left[-\frac{\tau}{\mu_o}\right] + \frac{F_T(\tau, \Omega)}{E(\tau)}\right] \exp\left[-\frac{\tau}{\mu}\right] \frac{d\tau}{\mu}\]
The total radiance reaching the detector is then the integral of $dI_D$ over all volume elements within $\Delta\omega$ between $z = -\infty$ and $z = z_o$ or, in terms of $\tau$, from $\tau = \infty$ to $\tau = 0$.

$$I_D = \int_{\tau=\infty}^{0} dI_D$$

$$= \int_{\tau=0}^{\infty} \left[ \frac{w(\tau)}{4\pi} \int_{4\pi} I(\tau, \Omega') p(\tau, g') d\Omega' + \frac{J}{4\pi} w(\tau)p(\tau, g) \exp \left[ -\tau/\mu_o \right] \right]$$

$$+ \left[ \frac{F_T(\tau, \Omega)}{E(\tau)} \right] \exp \left[ -\tau/\mu \right] \frac{d\tau}{\mu}$$

The bidirectional reflectance can then be written as

$$r = \frac{I_D}{J}$$

### 3.3.1 Single-scattering

If the contribution to $I_D$ is only singly-scattered light, then eq. (23) can be evaluated exactly. If the contribution of multiply-scattered light is negligible, then the first term in the integral in eq. (23) is zero. Moreover, if thermal emission is ignored, what remains is the singly-scattered term $I_{SS}$, which is given by

$$I_{SS} = \frac{J}{4\pi \mu} \int_{0}^{\infty} w(\tau)p(\tau, g)e^{-\tau(1/\mu_o+1/\mu)}d\tau$$

Invoking the fact that $w$ and $p$ are independent of $\tau$ in most applications, the integral can be evaluated to give

$$I_{SS} = J \frac{w}{4\pi \mu_o + \mu} p(g)$$
3.3.2 Multiple scattering for isotropic scatterers

The treatment of multiple-scattering in the Hapke model is based on the method of invariance. This method was first proposed by Ambartsumian (1958) and allows one to determine the reflectance and emittance of a semi-infinite particulate medium without solving the RTE. Instead, the invariance equation relies on the principle that the reflectance or thermal emission of an infinitely thick medium is not affected by the addition of a thin identical layer on top of it. This gives the following non-linear integral equation

\[
\frac{\mu_o + \mu}{\mu_o \mu} r(\Omega_o, \Omega) = \frac{\omega}{4\pi} \left[ \frac{1}{\mu} p(\Omega_o, \Omega) + \int_{\Omega_o} p(\Omega_o, \Omega') \frac{r(\Omega_o, \Omega)}{\mu_o'} d\Omega_o' \right. \\
\left. + \int_{\Omega'} p(\Omega', \Omega) \frac{r(\Omega_o, \Omega)}{\mu} d\Omega' \right. \\
\left. + \int_{\Omega_o'} \int_{\Omega_o'} p(\Omega', \Omega') r(\Omega_o, \Omega') \frac{r(\Omega_o', \Omega)}{\mu_o'} d\Omega_o' d\Omega' \right] \tag{26}
\]

where \( r(\Omega_o, \Omega) \) is the bidirectional reflectance for light coming from a source found in the direction \( \Omega_o = \Omega_o(i, \psi_i) \) that is scattered towards the detector in the direction \( \Omega = \Omega(e, \psi_e) \). This equation assumes that the optical thickness of the added layer can be made so small that its square and higher power terms can be ignored. This leaves only five terms describing interactions of light with the layer, each of which is proportional to the first power of the optical thickness.

A full derivation of eq. (26) can be found in Hapke (2012b). It is found by equating to zero the sum of the five first-order changes that occur due to the added layer. These changes are

1. Extinction of the incident light that occurs as it passes through the added layer on its way in, towards the medium, and on its way out, towards \( \Omega \). This change is described by the term on the left-hand side of eq. (26).
2. Scattering of the incoming light by the added layer towards the detector, which is described by the first term on the right-hand side.

3. Scattering of the incoming light by the added layer downward towards the medium. This light acts as an additional source of illumination to the medium and is described by the second term on the right-hand side, with \( \Omega'_o \) being integrated across the lower hemisphere.

4. Scattering of the incoming light by the medium upwards towards the added layer; some of this light is then scattered by the added layer towards the detector. The third term in eq. (26) describes this change, with \( \Omega' \) being integrated across the upper hemisphere.

5. Incident light is scattered by the medium, illuminating the added layer as in the above case; unlike the above case, however, the added layer scatters some of this light back down towards the medium, which in turn scatters some it towards the detector. The last term on the right-hand side of eq. (26) describes this, with \( \Omega' \) being integrated across the upper hemisphere and \( \Omega'_o \) being integrated across the lower hemisphere.

Letting

\[
  r(\Omega_o, \Omega) = \frac{w}{4\pi} \frac{\mu_o}{\mu + \mu_o} L(\Omega_o, \Omega) \tag{27a}
\]

and substituting it into eq. (26) gives

\[
  L(\Omega_o, \Omega) = p(\Omega_o, \Omega) + \frac{w}{4\pi} \mu_o \int_{\Omega'_o} p(\Omega_o, \Omega'_o) \frac{L(\Omega'_o, \Omega)}{\mu'_o + \mu} d\Omega'_o
  + \frac{w}{4\pi} \mu_o \int_{\Omega'} p(\Omega', \Omega) \frac{L(\Omega_o, \Omega')}{\mu_o + \mu'} d\Omega'
  + \left( \frac{w}{4\pi} \right)^2 \mu_o \mu \int_{\Omega'} \int_{\Omega'_o} p(\Omega', \Omega'_o) \frac{L(\Omega_o, \Omega')}{\mu_o + \mu'} \frac{L(\Omega'_o, \Omega)}{\mu'_o + \mu} d\Omega'_o d\Omega' \tag{27b}
\]

Ambartsumian (1958) showed that for the special case of isotropic scattering, when
\( p(g) = 1 \), \( L(\Omega_o, \Omega) \) is independent of azimuth and is only a function of \( \mu_o \) and \( \mu \), in which case, \( L(\Omega_o, \Omega) \) can be written as

\[
L(\mu_o, \mu) = \left[ 1 + \frac{\omega}{2} \mu_o \int_0^1 \frac{L(\mu_o, \mu')}{\mu_o + \mu'} d\mu' \right] \left[ 1 + \frac{\omega}{2} \mu \int_0^1 \frac{L(\mu'_o, \mu)}{\mu'_o + \mu} d\mu'_o \right]
\]

(28a)

where

\[
H(x) = 1 + \frac{\omega}{2} x H(x) \int_0^1 \frac{H(x')}{x + x'} dx'
\]

(28b)

is the Ambartsumian-Chandrasekhar \( H \) function \( \text{[Ambartsumian, 1958; Chandrasekhar, 1960]} \). Therefore, for the case of isotropic scattering, the exact solution of the bidirectional reflectance is

\[
r(i, e, g) = \frac{\omega}{4\pi} \frac{\mu_o}{\mu_o + \mu} H(\mu_o) H(\mu)
\]

(29)

Numerical solutions to the \( H \)-functions for isotropic scatterers have been published by Chandrasekhar (1960). However, in keeping with the theme of developing analytic solutions, Hapke (2002) derived an approximation given by

\[
H(x) \simeq \left( 1 - \omega x \left[ r_o + \frac{1 - 2r_o x}{2} \ln \left( \frac{1 + x}{x} \right) \right] \right)^{-1}
\]

(30a)

where

\[
r_o = \frac{1 - \gamma}{1 + \gamma}
\]

(30b)

This approximation has errors smaller than 1% everywhere when compared to exact solutions of Chandrasekhar (1960).
3.3.3 Anisotropic scattering

A widely used expression for the bidirectional reflectance in studying planetary regolith is what Hapke calls the isotropic multiple-scattering approximation (IMSA). This approximation relies on the fact that, for a medium that does not deviate too much from an isotropic scattering behavior, most of the anisotropic behavior is carried in the single scattering term. Chandrasekhar (1960) and Hansen and Travis (1974) have noted that the brighter a surface is, the more times a photon is scattered before emerging from the surface. This is a random process, so any directionality in the scattering is lost and the medium behaves more like an isotropic scatterer (Hapke, 1981, 2012b). This means the multiply scattered portion of light in bright surfaces is less sensitive to the particle phase function \( p(g) \), so the solution for isotropic scatterers can be used to approximate it, even if the medium consists of anisotropic scatterers. The exact solution for single scattering in a medium with arbitrary \( p(g) \) is given by eq. (25), so the multiply scattered portion of light for isotropic scatterers is found by subtracting eq. (25) from eq. (29) with \( p(g) = 1 \) (note that \( r = I/J \))

\[
I_{MS} = J \frac{w}{4\pi} \frac{\mu_o}{\mu_o + \mu} [H(\mu_o)H(\mu) - 1]
\]

Hence, the bidirectional reflectance can be approximated by

\[
r(i, e, g) = \frac{w}{4\pi} \frac{\mu_o}{\mu_o + \mu} [p(g) + H(\mu_o)H(\mu) - 1]
\]

(31)

This is the IMSA model that is frequently used to analyze photometric data from planetary surfaces.
3.4 Effects of increased filling factor

The derivations thus far were made under the assumption of sparse packing. However, planetary regolith can have varying degrees of packing, and the state of compaction has been shown to have an effect on the observed reflectance of particulate media in various studies, including Hapke and Wells (1981); Kar et al. (2016); Näränken et al. (2004); and Shepard and Helfenstein (2007). There are major changes that occur in the reflectance when particles are brought closer together: the Fraunhofer diffraction pattern vanishes, light transmissivity changes, and coherence effects become important.

3.4.1 Loss of Fraunhofer diffraction

The most noticeable change that occurs is the loss of the Fraunhofer diffraction pattern. The reflectance pattern of large (compared to the wavelength) isolated particles or sparsely packed particles has a sharp peak in the forward direction centered around $g = 180^\circ$. In addition, such particles have extinction efficiencies of about 2, not 1. Both of these phenomena arise because of the wave nature of light and cannot be explained by geometric optics. The source of this narrow peak is the constructive and destructive interference of the waves that pass by the particle. For spherical or approximately spherical particles, the interference pattern is virtually the same as that of an opaque disk. The diffraction pattern of such particles can thus be described by replacing this disk with the complementary scenario of an opaque wall with a hole of the same size and applying Babinet’s principle (Babinet, 1837). This analysis shows that the diffractive scattering efficiency $Q_d$ is unity, i.e., the power in the diffraction pattern, which mainly comes from light passing between the particle’s surface and a distance $\sqrt{2}a$, is equal to that intercepted by the cross sectional area of the particle $\sigma$. Adding this diffractive scattering to other scattering and absorption effects gives $Q_E \simeq 2$. Physically, what this means is that the particle affects a portion of
the wavefront that is about twice its cross sectional area.

When particles are brought closer together, however, the Fraunhofer diffraction pattern is altered because the waves passing by a particle are blocked by neighboring particles. This decreases the intensity of the diffracted light. Furthermore, with increased particle density, the diffracted light becomes highly concentrated in the forward direction that only the central (zeroth-order) peak remains, whilst other peaks vanish. The angular width of this central peak is of the order of the wavelength divided by the size of the medium. For planetary regolith, the medium size is of the order of the planet, hence the peak is so narrow that it is indistinguishable from the incident collimated light (Hapke, 1999). Therefore, a Fraunhofer diffraction pattern does not exist in a closely-packed medium.

Another phenomenon that must be addressed in close-packed media is Fresnel diffraction, which is important when the source and detector are close. The “source” in this case is a particle of interest from which light is scattered off of and a “detector” is a particle behind the source onto which the light falls. The light falling on the detectors is a Fresnel diffraction pattern, and calculating the exact pattern behind all of the irregularly shaped particles in a regolith is much too difficult a task. Therefore, approximations need to be made to describe multiply scattered component of the reflectance. It may be assumed that particles behind the source are randomly positioned so that they lie in bright and shadowed regions in proportion to the sizes of these regions. In addition, if the light passing in between particles, which normally has an intensity that depends on lateral distance from a particle, is assumed to be unscattered, then it can be averaged laterally and replaced by a uniform wavefront with intensity that is proportional to the fraction of open region. This means that \( Q_d = 0 \), so that for large particles \( Q_E \approx 1 \).
3.4.2 Transmissivity correction

The other change that occurs when particles are close together is in the transmissivity function. From the discussion in section [3.1.6] the intensity of light propagating in a continuous or sparsely-packed medium is assumed to attenuate exponentially (eq. [8]), with a *transmissivity function* $T$ that is given by

$$T = \exp \left( - \int_{s_1}^{s_2} E ds \right)$$

This function was derived with the assumption that the fraction of light intercepted by particles is negligible compared to the fraction that propagates unhindered, which is incorrect for closely-packed particles. Instead, the medium can be thought of as being made up of a lattice of cubes with side length $L$ with the center of a particle located inside each of these cubes. Each cube can contain only one particle so that there is no overlap between particles.

Now let’s consider the case of an incident light of radiance $I(s, \Omega)$ propagating perpendicularly through a slab of area $A >> L^2$ and thickness $L$. The slab contains $NAL$ particles each with extinction cross section $\sigma Q_E$. Then the power per unit solid angle of light that encounters particles in the slab and is extinguished is $dP_E/d\Omega = I(NAL)\sigma Q_E = IAEL$. The amount that is transmitted is $dP_T/d\Omega = IA - IAEL = IA(1 - EL)$. Hence, the transmissivity function is the ratio of the transmitted power to the incident power per solid angle $IA(1 - EL)/IA = (1 - EL)$. Assuming that the particles are randomly positioned inside each slab, the transmissivity function for light propagating through several layers is simply the product of the transmission probability through each layer. So for $n = s/L$ layers,

$$T(s) = (1 - EL)^n = \exp[n \ln(1 - EL)] = \exp(-KEs) \quad (32a)$$
where

$$K = -\ln(1 - EL) / EL$$

(32b)

is the porosity coefficient. The above expression for $T(s)$ is a step function (Fig. 7) that says incident light is attenuated by a factor $(1 - EL)^n$ after propagating a distance $s$ where $n \leq s/L \leq (n + 1)$. The value of $K \to 1$ when particles are far apart, but $K > 1$ for other cases. This means light attenuates more sharply in close-packed media than in a continuous media. However, the upper layers of the medium receive more light than would a continuous medium (Fig. 7).

Figure 7: Transmissivity functions when $EL = 0.4$ for a sparsely-packed continuous medium (dotted curve) and for discrete (solid line) and continuous (dashed curve) layers of closely-packed media, from Hapke (2008).

For equant particles, i.e. those having roughly the same dimensions in all directions,
\[ EL \approx 1.209\phi^{2/3} \] so that

\[
K = -\frac{\ln(1 - 1.209\phi^{2/3})}{1.209\phi^{2/3}}
\] (33)

The medium becomes opaque when \( EL = 1 \) or when \( \phi = 0.752 \). However, before this limit is reached coherent effects become important.

### 3.4.3 Coherence effects

When the separation distance between particles becomes small, effects produced by inter-particle interaction become important. The region near the surface of an illuminated particle not only has waves propagating radially outward with intensity that falls off as the square of distance from the particle center, but also near fields which are waves that fall off faster. Moreover, when the conditions for total internal reflection are met at the particle-vacuum interface, a system of moving fields called evanescent waves actually penetrate a short distance into the vacuum medium. In a media of low filling factor, these waves do not carry energy away from the particle surface. However, if the particle separation is smaller than a wavelength, these waves will start to induce charges and currents in neighboring particles leading to a transfer of energy between particles. This is a form of scattering that cannot be ignored. At low filling factors, particles also start to clump together and act as a single, larger particle. For a medium of equant particles, the threshold beyond which coherent effects become important is

\[
\phi > \frac{\pi/6}{(1 + \lambda/D)^{3/4}}
\] (34)

As shown in Fig. 8 for a particle large compared to the wavelength, coherence effects are important when \( \phi \gtrsim 50\% \) while it is significant for all close-packed particles with size of the order of a wavelength or smaller. Below the coherence curve, the medium can be
thought of as a continuous void with some solid particles and above the curve the medium is a nearly continuous solid with some empty spaces scattered in it. The Hapke reflectance model is applicable only in the region where coherence effects are negligible, hence it does not apply to compacted solids.

3.4.4 Bidirectional reflectance at arbitrary filling factor

The RTE given in section 3.2 is correct only for sparsely-packed media. For arbitrary filling factors, the RTE must be modified to take into account the changes that occur when particles are placed close to each other. Consider a radiance \( I(z, \Omega) \) propagating at a depth \( z \) through a layer of closely-packed particles with thickness \( \Delta s = L \) making an angle \( \epsilon \) with the vertical in a direction \( \Omega \). If thermal emission is neglected, the change in radiance as the light propagates through the layer is

![Figure 8: Plot of \( \phi \) vs. \( D/\lambda \) showing the limits of when coherence effects become important and when the medium becomes opaque to incident light, from Hapke (2008).](image)
\[ \Delta I(z, \Omega) = -ELI + \frac{SL}{4\pi} \int_{4\pi} I(z, \Omega)p(g')d\Omega' + J\frac{SL}{4\pi}p(g)T(z/\mu_o) \]

where \( T(s) = T(z/\mu_o) \) is the transmissivity function given by eq. (32a) with \( \mu_o = \cos(i) \).

This equation is similar to equation eq. (20) except for the fact that it is discontinuous. It can be approximated by a continuous function by making two assumptions: (1) the change in radiance varies linearly through a layer so that \( \Delta I \) can be written as

\[ \Delta I(z, \Omega) = \frac{\Delta I(z, \Omega)}{\Delta s} L \approx \frac{\partial I(z, \Omega)}{\partial s} L = \mu \frac{\partial I(z, \Omega)}{\partial z} L = -\mu \frac{\partial I(z, \Omega)}{\partial \tau} EL \]

where \( \mu = \cos(e) \) (2) the discontinuous function of eq. (32a) can be substituted by the continuous function \( C \exp(-KEs) \), where \( C \) is a constant determined by the limiting condition that the radiance \( I(0)(1 - EL)'^n \) incident on a particle between \( s = nL \) and \( s = (n + 1)L \) in the discontinuous medium is equal to the average radiance in the continuous medium between these distances, which is given by

\[ I(0)\frac{C}{L} \int_{nL}^{(n+1)L} \exp(-KEs)ds. \]

This leads to the result \( C = K \). The transmissivity function thus becomes

\[ T(s) = K \exp(-KEs) \quad (35) \]

It can be seen from Fig. [7] that this equation passes through the mid-point of each step.

With the above approximations, the RTE takes the form of eq. (22) for a sparse media, except for the factor of \( K \) in the transmissivity function and the lack of Fraunhofer diffraction in the efficiencies

\[ -\mu \frac{\partial I(\tau, \Omega)}{\partial \tau} = -I(\tau, \Omega) + \frac{w}{4\pi} \int_{4\pi} I(\tau, \Omega')p(g')d\Omega' + J\frac{w}{4\pi}p(g)K \exp[-\tau/\mu_o] \quad (36) \]
The reflectance equation that follows from this RTE is

\[ r(i, e, g) = K \frac{\omega}{4\pi} \frac{\mu_o}{\mu + \mu} [p(g) + H(\mu_o/K)H(\mu/K) - 1] \tag{37} \]

### 3.5 The opposition effect

The other major change in the scattering of light by particulate media that occurs as particles are brought together is the opposition effect. This is a ubiquitous phenomenon characterized by a surge in the observed reflectance as the phase angle approaches zero. It was first observed by Seeliger (1887, 1895) in light scattered from Saturn’s rings, and since then Seeliger and others have proposed several mechanisms to explain the opposition effect. However, in the most recent interpretation, the opposition effect is attributed to two causes: shadow hiding effect and coherent backscattering. The form of Hapke model used in the present study does not include the coherent backscattering opposition effect, so only the shadow hiding effect is discussed below. Given the absence of data near opposition in this study \((g \geq 14^\circ)\), these effects are thought to contribute little to the observed reflectance. In addition, coherent backscattering is believed to become important at phase angles \(\leq 3^\circ\) (Hapke et al., 1998), so not taking its effect into account should not have any significant influence on the results.

#### 3.5.1 Shadow hiding opposition effect

The shadow hiding opposition effect (SHOE) occurs in particles that are large enough to have shadows. This occurs in media with particles larger than the illuminating wavelength. The shadows are visible at large phase angles but are blocked by the particle casting the shadow at small phase angles. Another way of understanding this phenomenon is to think of the spaces between particles as tunnels through which light can travel without being
extinguished. At large phase angles, the tunnel is blocked by particles making up its walls, but at small phase angles the path taken by the light is visible, so the particles illuminated at the bottom of the tunnel are also visible. This results in an increase in the observed reflectance.

The RTE fails to account for the SHOE, so it must be added onto the Hapke reflectance equation. As it will be discussed in later paragraphs, the effect is important only in singly-scattered light, so only that portion of the RTE needs to be modified. From sections 3.3.1 and 3.4.2 the singly-scattered radiance is

\[ I_{SS} = \frac{J^w}{4\pi} \mu p(g) \int_0^\infty T_i(\tau, \mu_o) T_e(\tau, \mu) d\tau \]

where

\[ T_i(\tau, \mu_o) = K \exp(-K\tau/\mu_o) \]

is the probability for the incident light to reach a depth \( z \) where the optical depth is \( \tau = \int_z^\infty E(z')dz' \) and

\[ T_i(\tau, \mu_o) = K \exp(-K\tau/\mu) \]

is the probability that scattered light will reach the top of the surface from a depth \( z \). According to the RTE, these probabilities are independent of each other, so the total probability is just their product and the integration result is given by eq. (25). This is true at large phase angles, but is incorrect for small phase angles. To see why, let’s consider Fig. 9 in which a ray of incident light \( J \) is scattered at point \( P \) and exits the surface as radiance \( I \).

Following from eq. (14a), let \( \sigma_E = \sigma Q_E = E(z)/N(z) \) be the volume averaged extinction cross section and \( a_E = \sqrt{\sigma_E/\pi} \). Now an equivalent definition for \( T_i \) to the one given
above is that it is the probability that no particle has its center in a cylinder with radius $a_E$ and is coaxial with the incident ray; similarly, $T_e$ is the probability that no particle has its center in the cylinder with the same radius but that is coaxial with the scattered light. At small phase angles, these cylinders start to overlap, so a total probability calculated as $T_i T_e$ double counts the overlap region. To correct this, the total probability needs to be redefined as

$$T_i T_e = K^2 \exp[-K(\tau/\mu_o + \tau/\mu - \tau_c)]$$  \hfill (38)

where $\tau_c = \int_{V_c} N(z')dV$ and $V_c$ is the volume of the overlap region whose cross section is APBC in Fig. 9. The value of $\tau_c$ is trivial when $g = 0$; in this case, the cylinders overlap

Figure 9: Schematic diagram showing source of shadow hiding effect in singly-scattered light, from [Hapke (2012b)].
perfectly, so \( \tau_c = \tau/\mu = \tau/\mu_o \), and

\[
I_{SS}(i, e = i, 0) = \frac{J w}{4\pi} \mu p(0) \int_0^\infty K^2 \exp[-K\tau/\mu]d\tau = \frac{J}{4\pi} K \mu p(0)
\]

This result is exactly twice of that predicted by the RTE. Therefore, the correlation of the transmissivity functions says that if light propagates to an optical depth \( \tau \) without being extinguished, then the probability that it travels the same path in the opposite direction without experiencing any attenuation is one, not \( \exp(-\tau/\mu) \).

The general case of a non-zero phase angle is more involved, but [Hapke (1986, 2012b)] gives an analytic solution by making some geometric and small-angle approximations to find the common volume of the intersecting cylinders. This approximation is then used to calculate \( \tau_c \). The details of the derivation are described in the original references and lead to an expression for the singly-scattered radiance given by

\[
I_{SS} = J \frac{w}{4\pi} \frac{\mu_o}{\mu_o + \mu} K p(g) \left[ 1 + B_{S0} B_S(g) \right] \tag{39a}
\]

where

\[
B_S(g) = \left[ 1 + \frac{1}{h_S} \tan \left( g/2 \right) \right]^{-1} \tag{39b}
\]

is the shadow hiding opposition function with the property \( B_S(0) = 1 \) and an angular half-width (HWHM) that occurs when \( (1/h_s) \tan(g/2) = 1 \) or approximately when \( g = 2h_S \).

For a semi-infinite medium with a constant particle density distribution, the angular width parameter can be written as

\[
h_S = \frac{KEa_E}{2} \tag{39c}
\]

This means that as \( \phi \) increases, the angular width increases and becomes infinitely wide.
when $\phi = 0.752$, i.e., when the medium becomes opaque. Hence, opaque particles will not have a SHOE. At the other extreme, as $\phi$ decreases the angular width decreases and the peak becomes narrow; however, it can’t be narrower than the angular widths of the source and detector and has a lower limit of the sum of the angular half-widths of the source and detector. The coefficient $B_{S0}$ is the amplitude of the SHOE and is in the range $0 \leq B_{S0} \leq 1$. It is an empirical factor added to account for the finite size of particles.

Incident light illuminating a particle may scatter and emerge from a different side of the particle than its entry point in which case the overlap region of the probability cylinders is reduced. The amplitude term is interpreted as a measure of transparency; it is the ratio of light scattered at or close to the front surface (illuminated part facing the source) of the particle to the total amount of light scattered by the particle at zero phase:

$$B_{S0} \approx \frac{S(0)}{wp(0)}$$  \hspace{1cm} (39d)

It approaches one when light is backscattered at the surface or after traversing a small distance into the particle.

Multiple scattering has little contribution to the SHOE because the probability of overlap between the incident and emergent cylinders is negligible. Esposito (1979) has shown that inclusion of doubly-scattered light makes an insignificant difference to the reflectance. Therefore, it is adequate to multiply the shadow hiding function only to the single scattering term in the reflectance function.

### 3.6 Macroscopic Roughness

The treatment of the bidirectional reflectance thus far does not take into account variations in surface topography. Although laboratory samples may be assumed to be smooth, planetary regolith posses varying degrees of roughness which affect the predicted reflectance.
Hapke (1984) derived an expression to correct the bidirectional reflectance of a smooth surface to include surface roughness that is characterized by the mean slope angle $\overline{\theta}$ parameter. The derivation remains analytic, albeit makes some mathematical approximations and physical assumptions. The underlying assumptions are outlined as follows (Hapke, 2012b):

1. The model is valid within the domain of geometric optics, so the scattering elements are large compared to the wavelength. If the medium is composed of small particles, then the scattering elements are large clumps rather than individual particles.

2. The surface is made up of smooth facets that are larger than the mean particle size. Their orientation is described by a slope distribution function $a(\theta, \zeta)d\theta d\zeta$, where $\theta$ is the angle between the facet normal and the vertical and $\zeta$ is the azimuth angle of the facet normal.

3. The facet orientations are assumed to be independent of azimuth, so that $a(\theta, \zeta) = a(\theta)$. This assumption is valid on average for surfaces covered with craters, hills and boulders, but may be questionable for surfaces made up of ridges or dunes. However, it is important to note that in planetary applications, the relevant scale of roughness ranges from millimeters to kilometers. The lower bound of this scale is determined by the RTE in particulate media, which averages quantities over scales much larger than particle separation (section 3.1.6), while the upper bound is determined by the angular resolution of the detector. In addition to the local acceleration due to gravity, the largest slope on a surface depends on the strength of the material and the cohesive nature of the soil. The effects of these properties are size-dependent such that the largest slopes, which dominate the reflectance, occur at the lower end of the relevant size range. Surface roughening agents such as eolian gusting, fluvial actions, and microcratering are roughly independent of azimuth at the dominant small scale, hence, the assumption is reasonable.
4. The mean slope \( \bar{\theta} \) is assumed to be small. This assumption is made so that orders of \( \bar{\theta}^3 \) and higher can be ignored in the derivation, which simplifies the resulting expressions. The final equations are, however, valid for any arbitrary \( \bar{\theta} \).

5. The slope distribution function can be described by a Gaussian function. Let \( a'(\theta) \) represent the one-dimensional slope distribution in a vertical cut through the surface at an arbitrary azimuth. Then it is related to the corresponding two-dimensional, azimuth-independent distribution function by the equation [Hagfors, 1968]

\[
a(\theta)d\theta d\zeta = a'(\theta) \sin \theta d\theta d\zeta
\]

If \( a'(\theta) \) is assumed to be of the form

\[
a'(\theta)d\theta = A \exp\left[-B \tan^2 \theta\right]d(\tan \theta)
\]

then

\[
a(\theta) = A \exp\left[-B \tan^2 \theta\right] \sec^2 \theta \sin \theta
\]

where \( A \) and \( B \) are constants. The slope distribution function is normalized so that

\[
\int_0^{\pi/2} a(\theta)d\theta = 1
\]

and is characterized by \( \bar{\theta} \) which is given by

\[
\tan \bar{\theta} = \frac{2}{\pi} \int_0^{\pi/2} a(\theta) \tan \theta d\theta
\]

Substituting eq. (40b) into eqs. (41a) and (41b) and and using the assumption of small
\( \overline{\theta} \), the constants become

\[
A = \frac{2}{\pi \tan^{2} \overline{\theta}} \quad (41c)
\]

\[
B = \frac{1}{\pi \tan^{2} \overline{\theta}} \quad (41d)
\]

6. Inter-facet scattering is neglected, although multiply scattered light from one particle to another within a facet is taken into account in the derivation. As shown in Hapke (2012b), light multiply scattered from one facet to another has little contribution to the reflectance when the albedo or the mean slope is small. Hence, this assumption is consistent with assumption (4). However, it is not true for high-albedo surfaces at large phase angles. Inter-facet scattering in such surfaces is significant enough that areas that are shadowed from direct light from the source will still be illuminated due to multiply-scattered light from other facets. Therefore, shadows will be filled in making the surface photometrically smoother. This has the effect of lowering the photometric albedo. An empirical correction can be applied for high-albedo surfaces to account for the shortcoming caused by this assumption in the prediction of surface roughness.

The Hapke (1984) derivation corrects for roughness effects by writing the reflectance of a rough surface \( r_R(i, e, g) \) characterized by a mean slope \( \overline{\theta} \) as the product of a shadowing function \( S(i, e, g) \) and the reflectance of a smooth surface \( r(i_e, e_e, g) \) of effective area \( A_e \) with effective angles of incidence and emergence \( i_e \) and \( e_e \), respectively (the phase angle remains unchanged). Exact expression for the effective angles are found at vertical and grazing illumination and viewing. At intermediate angles, analytic interpolation is used to find approximate solutions. The complete derivation is involved, so will not be discussed here; instead, the results will be summarized.
The solutions have different forms depending on whether \( i \) is larger or smaller than \( e \).

For the case where \( i < e \)

\[
\mu_{0e}(i, e, \psi) \equiv \cos i_e = \chi(\bar{\theta}) \left[ \cos(i) + \sin(i) \tan(\bar{\theta}) \frac{\cos(\psi)E_2(e) + \sin^2(\psi/2)E_2(i)}{2 - E_1(e) - (\psi/\pi)E_1(i)} \right] \quad (42a)
\]

\[
\mu_{e}(i, e, \psi) \equiv \cos e_e = \chi(\bar{\theta}) \left[ \cos(e) + \sin(e) \tan(\bar{\theta}) \frac{E_2(e) - \sin^2(\psi/2)E_2(e)}{2 - E_1(e) - (\psi/\pi)E_1(e)} \right] \quad (42b)
\]

\[
S(i, e, \psi) = \frac{\mu_e}{\eta_e(e)} \frac{\mu_0}{\eta_{0e}(i)} \frac{\chi(\bar{\theta})}{1 - f(\psi) + f(\psi)\chi(\bar{\theta})[\frac{\mu_e}{\eta_e(i)}]} \quad (42c)
\]

and when \( i > e \)

\[
\mu_{0e}(i, e, \psi) \equiv \cos i_e = \chi(\bar{\theta}) \left[ \cos(i) + \sin(i) \tan(\bar{\theta}) \frac{E_2(i) + \sin^2(\psi/2)E_2(e)}{2 - E_1(i) - (\psi/\pi)E_1(e)} \right] \quad (43a)
\]

\[
\mu_{e}(i, e, \psi) \equiv \cos e_e = \chi(\bar{\theta}) \left[ \cos(e) + \sin(e) \tan(\bar{\theta}) \frac{\cos(\psi)E_2(i) + \sin^2(\psi/2)E_2(e)}{2 - E_1(i) - (\psi/\pi)E_1(e)} \right] \quad (43b)
\]

\[
S(i, e, \psi) = \frac{\mu_e}{\eta_e(e)} \frac{\mu_0}{\eta_{0e}(i)} \frac{\chi(\bar{\theta})}{1 - f(\psi) + f(\psi)\chi(\bar{\theta})[\frac{\mu_e}{\eta_e(i)}]} \quad (43c)
\]

The terms \( \chi, E_1, E_2, \eta_{0e}, \eta_e, \) and \( f \) are given by

\[
\chi(\bar{\theta}) = \frac{1}{\sqrt{1 + \pi \tan^2(\bar{\theta})}} \quad (44)
\]
\[ E_1(x) = \exp \left[ -\frac{2}{\pi} \cot(\theta) \cot(x) \right] \] (45)

\[ E_2(x) = \exp \left[ -\frac{1}{\pi} \cot^2(\theta) \cot^2(x) \right] \] (46)

\[ \eta_{oe}(i) = \chi(\bar{\theta}) \left[ \cos(i) + \sin(i) \tan(\bar{\theta}) \frac{E_2(i)}{2 - E_1(i)} \right] \] (47)

\[ \eta_{e}(e) = \chi(\bar{\theta}) \left[ \cos(e) + \sin(e) \tan(\bar{\theta}) \frac{E_2(e)}{2 - E_1(e)} \right] \] (48)

\[ f(\psi) = \exp \left[ -2 \tan \frac{\psi}{2} \right] \] (49a)

where \( \psi \) is the azimuth angle (Fig. 6) and is related to the incident, emergence, and phase angles by

\[ \cos(\psi) = \frac{\cos(g) - \cos(i) \cos(e)}{\sin(i) \sin(e)} \] (49b)

Combining the porosity coefficient, SHOE, and surface roughness terms, the expression for the bidirectional reflectance becomes

\[ r(i, e, g) = K \frac{w}{4\pi} \frac{\mu_{oe}}{\mu_{oe} + \mu_e} \{ p(g)[1 + B_{SO}B_S(g)] + H(\mu_{oe}/K)H(\mu_e/K - 1) \} S(i, e, \psi) \] (50)
4 ISS Instrument and Image Calibration

4.1 Mission Overview

The Cassini/Huygens mission was a joint project among the National Aeronautics and Space Administration (NASA), the European Space Agency, and the Italian Space Agency to “conduct an in-depth exploration of the Saturnian System” (NASA 1989). It consisted of the Cassini Saturn Orbiter spacecraft, for studying the whole system, and the Huygens Titan Probe, for entry and descent into Saturn’s planet-sized moon Titan. The science goals of the mission were extensive with specific objectives for each type of body in the Saturn system, i.e., the planet, satellites, rings, and magnetosphere. For icy satellites, some of the objectives were

- determine physical properties, geology and surface and internal processes
- map composition of surface materials
- examine interactions with rings and magnetosphere

But beyond this, Cassini/Huygens was designed to study the interactions that occur between and among these bodies. Since many of the phenomena to be measured depended on several parameters, it was important to have the right types of instruments that could take simultaneous measurements of their surrounding. To this end, Cassini/Huygens was equipped with a suite of in-situ and remote sensing instruments, twelve on the Orbiter and six on the Huygens Probe (Matson et al. 2002).

One of the remote sensing instruments onboard the Cassini Orbiter was the Imaging Science Subsystem (hereafter ISS). The ISS was the highest resolution two-dimensional imaging device that was designed for imaging bodies and phenomena within the Saturn system (Porco et al. 2004). The following sections provide a description of the Cassini...
ISS. A more detailed accounting of instrument characteristics for the ISS can be found in Porco et al. (2004).

### 4.2 Instrument Characteristics

The ISS consists of two fixed focal-length cameras attached to the Remote Sensing Palette of the *Orbiter* that are bore-sight aligned. The narrow-angle camera (NAC) is an $f/10.5$ reflecting telescope with an image scale of $\sim 6\mu$ rad/pixel and a $0.35^\circ \times 0.35^\circ$ field-of-view (FOV). Its spectral sensitivity ranges from 200 to 1100 nm with its filter wheel subassembly carrying 24 spectral filters that are mounted on two wheels, each carrying 12 filters. The wide-angle camera (WAC) is a spare Voyager-era refracting telescope with an $f/3.5$ aperture, an image scale of $\sim 60\mu$ rad/pixel, and a $3.5^\circ \times 3.5^\circ$ FOV. Due to its refractor design, the spectral sensitivity of the WAC is limited to 380 to 1100 nm with nine filters in each of its two filter wheels. The cameras differ in their optical design and number of filters but are otherwise similar. Each has its own set of optics, charge-coupled device (CCD), shutter, filter wheel, temperature sensors, heaters, control electronics, flight computer, and Bus Interface Unit to the spacecraft’s Command and Data System (Porco et al. 2004, Fig. 10).

### 4.3 Filters

The selection of filters in both cameras was determined by the science objectives, instrument capabilities as well as the nature of the flight paths of the spacecraft. A mission like *Cassini*, studying different targets at varying distances, makes highly-eccentric orbits and fast flybys. During closest approach of a target, large variation in viewing geometry and phase angle occurs that may not be repeated over the course of the mission. On approach, the NAC makes spectrophotometric observations but at closest approach, it does not have
Figure 10: Photographs of flight models of (a) NAC Head Assembly and (b) WAC Head Assembly from Porco et al. (2004).
enough time to cover the required spatial extent due to its smaller FOV. Hence, at such
times, similar observations are made using the WAC. To this end, more than half of the
filters in the NAC are duplicated in the WAC. These include seven medium/broad-band fil-
ters ranging from blue to near-IR, two methane and two continuum band filters, two clear
filters, and a narrow-band H\(\alpha\) filter (Porco et al., 2004).

The filter assembly of each camera consists of two overlapping filter wheels and a filter
changing mechanism. Each wheel can rotate independently of the other wheel, giving flex-
ibility in filter combination. The different types of filters in both cameras are summarized
here from Porco et al. (2004).

- Each camera has clear filters that allow transmission across the entire spectral extent
  of the instrument. These filters are particularly useful for imaging faint objects and
  for increasing signal-to-noise ratio when short exposure times are required to mini-
mize smear. They are typically used together with a color filter or a polarizer found
  on the other wheel.

- There are also sets of medium and broad-band filters in each camera that cover the
  spectral range of the CCD. They include BL1, GRN, RED, IR1, IR2, IR3, and IR4
  filters found on both cameras; UV1, UV2, and UV3 filters in the NAC; and VIO and
  IR5 filters in the WAC. The NAC has sensitivity in the UV range due to its lumogen
  coating, a capability that was unavailable during the Voyager and Galileo missions.

- Atmospheric studies necessitated the use of narrow band filters. Methane absorption
  band and continuum filters MT1 and CB1 in the NAC and MT2, CB2, MT3, and
  CB3 in both cameras were used to see through Titan’s atmosphere. The HAL filter
  allows for observation of H\(\alpha\) emissions from Saturn’s lightning.

- Polarizing filters are also included for studying the scattering properties of atmo-
spheres, rings, and satellite surfaces. The NAC has three polarizers in the visible spectrum— P0, P60, and P120— and one polarizer, IRP0, in the infrared. The WAC includes only two infrared polarizers, IRP0 and IRP90, due to the limited number of filter slots available.

4.4 Shutter

The ISS shutter assembly is a two-blade mechanical system situated between the filter wheel assembly and the CCD detector. The blades move independently of each other in the sample (row) direction of the CCD. The shutter operates in three phases: 1) in the open phase one blade moves in the negative sample direction, i.e. from the last row to the first, to open the shutter 2) in the close phase the second blade follows the first to close the shutter 3) in reset phase both blades move back to their start positions for the next exposure (Porco et al., 2004).

There are 63 commanded exposure settings ranging from 5ms to 1200 seconds and one “No-Op” (no operation) setting in which no shutter movement occurs. Due to mechanical imperfections of the assembly, there are offsets in actual exposure times compared to the commanded times. The mean offsets for the NAC and WAC are 2.85 ms and 2.86 ms less than the commanded time, respectively, with an uncertainty of ± 0.25 ms (Knowles, 2018). There is also an offset that occurs due to movement of the shutter blades in the sample direction. Because the last column of the CCD is illuminated first, its actual exposure time is longer than that of the first column’s. In the NAC, it is 0.3 ms longer and in the WAC it is 0.1 ms longer. Corrections for both of these shutter offsets are incorporated in the exposure correction algorithm of the calibration software (Porco et al., 2004).
4.5 Detector

Both ISS cameras consist of a three-phase, front-side illuminated CCD detector with 1024 × 1024 pixels each 12 µm on a side. The response of each camera’s CCD to incident photons is determined by the respective QE curve (Fig. 11). The curve approximately stays constant at 0.13 e⁻/photon from 200 to 450 nm but then increases up to 550 nm. The QE of both the NAC and WAC peaks in the wavelength range 550-800 nm before falling steeply in the near-IR region and reaching zero at 1100 nm. Although the WAC CCD is sensitive to UV light, its Voyager era optics are opaque at this wavelength, limiting its spectral capability (Porco et al., 2004).

Figure 11: Quantum efficiency curve (in e⁻/photon) of NAC and WAC as measured during ground calibration of the ISS from [Porco et al. (2004)]. Note that these values do not include QE correction from in-flight calibration (see section 4.7).

The CCD can be operated in “full” mode, corresponding to no binning, or in 2×2 or 4×4 summation modes. Summation modes are used to increase signal-to-noise ratio, de-
Gain Index | Description | GAIN_MODE_ID | NAC Gain (e'/DN) | Ratio (g_e/g_i) | WAC Gain (e'/DN) | Ratio (g_e/g_i)
---|---|---|---|---|---|---
0 | Designed for 4×4 | ‘215 ELECTRONS PER DN’ | | 0.135 | 0.125 |
1 | Designed for 2×2 | ‘95 ELECTRONS PER DN’ | | 0.310 | 0.291 |
2 | Used in 1×1 | ‘29 ELECTRONS PER DN’ | 30.27 | 1.000 | 27.68 | 1.000 |
3 | Used in 1×1 | ‘12 ELECTRONS PER DN’ | 2.357 | 2.360 |

Table 1: Gain states of the ISS from Knowles (2018). Gain state 3 was chosen to match the CCD read noise.

crease data volume, or decrease readout time. There are four gain states available with these summation modes: gain states 0, 1, and 2 correspond to 4×4, 2×2, and 1×1 summations, respectively, while the gain state 3 is a high gain used in 1×1 mode for imaging faint objects (Table 1). The CCD pixel full-well is approximately 120,000 electrons and the summation-well capacity, corresponding to full-well for 4×4 summing, is ~1.6×10^6 electrons—summation-well capacity does not simply scale with the number of binned pixels. However, non-linearity of CCD response starts to become an issue for 4×4 summation at levels of 10^6 electrons. For this reason, the largest possible DN output of the detector was set to this signal level (Porco et al., 2004).

### 4.6 Data Conversion and Compression

The amount of data that can be collected by the ISS is critically determined by the limited storage available in each camera’s Solid State Recorder (SSR) and the limited bandwidth available for communication back to Earth. Therefore, it is important to perform data acquisition and transfer in an efficient manner. Three camera settings afford this capability: on-chip binning—discussed above—data conversion, and data compression.
4.6.1 Data Conversion

During readout of the CCD, charges in the serial register are passed through the output amplifier and digitized by the A/D converter. In ISS, data are digitized to 12-bit DN values with a dynamic range of 4096. However, they are stored in the spacecraft in 16-bit format: the upper four bits are set to 1’s and get masked by the flight software during calculations. After digitization, there is the option to perform data conversion, which involves converting the 12-bit values to 8-bits, or keep the values as is. Data conversion can be carried out in one of two ways: (1) in the first method, the 12-bit values are simply truncated to the least-significant 8-bit values (LS8B). This type of conversion is most useful for reducing data volume of faint scenes whose DN values do not exceed 255; (2) in the second method, a variant of the square-root encoding is used. This encoding uses a look-up table (LUT) to convert 12-bit values to 8-bit by taking the square-root of the 12-bit input (Fig. 12). The advantage of this method is that it closely matches the quantization-level to the photon shot noise, which is a noise that arises from variation in the amount of photons hitting the detector, enabling information to be spread more evenly across the 256 DN values (Porco et al., 2004).

4.6.2 Data Compression

The next processing step is to decide what kind of compression the data goes through. Unconverted images can undergo lossless compression or no compression while converted images can undergo lossless, lossy, or no compression. The ISS lossless compression algorithm uses Huffman encoding (Huffman 1952) to compress data. In this scheme, the length of the bit sequence used to encode a number is chosen based on the frequency of occurrence of the number. Lossless compression, as its name suggests, compresses data with no loss of information provided that the image compression ratio is no less than the 2:1
threshold. The highest compression ratio that can be achieved for an image is determined by its entropy, which is a measure of the amount of visual information contained in an image. Scenes with low image entropy can be compressed to more than 2:1 ratio while high image entropy scenes will never compress greater than 2:1. For images whose entropy is too high to achieve the minimum 2:1 lossless compression, the ISS compression algorithm truncates the ends of lines (rows) of pixels so that the information in a pair of lines does not exceed that contained in one uncompressed line. Truncation is done on every other line on the right-side of the image (Porco et al. 2004; Fig. 13).

The ISS lossy compression algorithm is a variation of the common Joint Photographic Experts Group (JPEG) algorithm. It removes high spatial frequency information while retaining important details. It is used on 8-bit converted images and can achieve a more stringent (high) compression ratio but is not used on images taken for photometry. Lossy compression was discontinued since the start of the Solstice Mission to save time spent on post-processing of images.
4.7 Radiometric Calibration

In this section, the theoretical basis for radiometric calibration will be discussed. The information presented here and in the next section is taken from [Porco et al. (2004)] and [Knowles (2018)], supplemented with descriptions from Cassini ISS Ground Calibration Report (located in the documents/report directory), which can be found on the PDS Cassini ISS online data volumes page:

https://pds-imaging.jpl.nasa.gov/volumes/iss.html

The goal of calibration is to relate measured DN values to a useful physical quantity for
scientific analyses, namely incident intensity, $I$. The definitions used in presenting the
theory are given in Table 2:

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Units</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Omega$</td>
<td>steradian</td>
<td>solid angle sampled by one pixel</td>
</tr>
<tr>
<td>$A$</td>
<td>cm$^2$</td>
<td>collecting area of camera optics ($0.25\pi d^2$; where $d$ is the diameter of lens/primary mirror)</td>
</tr>
<tr>
<td>$C(f_1, f_2)$</td>
<td></td>
<td>absolute sensitivity correction factor determined from in-flight calibration</td>
</tr>
<tr>
<td>$e_p(i, j)$</td>
<td>electrons</td>
<td>photoelectrons collected by CCD</td>
</tr>
<tr>
<td>$f_1, f_2$</td>
<td></td>
<td>filter in wheel 1 and filter in wheel 2</td>
</tr>
<tr>
<td>$FF(i, j, f_1, f_2)$</td>
<td></td>
<td>flat field relative sensitivity</td>
</tr>
<tr>
<td>$g$</td>
<td>electrons/DN</td>
<td>gain value (Table 1)</td>
</tr>
<tr>
<td>$I(i, j, \lambda)$</td>
<td>Photons/(cm$^2$ s nm steradian)</td>
<td>intensity at pixel $(i,j)$ and $\lambda$</td>
</tr>
<tr>
<td>$QE(i, j, \lambda)$</td>
<td>Electrons/photon</td>
<td>CCD quantum efficiency</td>
</tr>
<tr>
<td>$RBI(i, j, mode)$</td>
<td>DN</td>
<td>Residual Bulk Image sample</td>
</tr>
<tr>
<td>$t(i)$</td>
<td>seconds</td>
<td>sample dependent exposure time</td>
</tr>
<tr>
<td>$T_o(i, j, \lambda)$</td>
<td></td>
<td>optics transmission coefficient; accounts for beam obscuration as well as losses at lens and mirror surfaces</td>
</tr>
<tr>
<td>$T_1(i, j, \lambda)$</td>
<td></td>
<td>filter 1 transmission coefficient</td>
</tr>
<tr>
<td>$T_2(i, j, \lambda)$</td>
<td></td>
<td>filter 2 transmission coefficient</td>
</tr>
</tbody>
</table>

Table 2: Definitions of calibration quantities and units

Before light hits the CCD, it travels through the camera optics, two filters and the shutter
getting diminished by the transmission coefficients and characteristics of the components.
Hence, the number of photons that reaches a particular pixel $(i, j)$ on the detector is given
by \( A \Omega t(i) I(i, j, \lambda) T_o(i, j, \lambda) T_1(i, j, \lambda) T_2(i, j, \lambda) \). The exposure time is sample dependent because the shutter velocity is not uniform. After light hits the CCD, a fraction \( Q E(i, j, \lambda) \) of the photons are absorbed while the rest scatter without being detected. Therefore, the expression for the number of electrons that get released in pixel \((i, j)\) due to the photoelectric effect is

\[
e_p(i, j) = A \Omega t(i) \int I(i, j, \lambda) T_o(i, j, \lambda) T_1(i, j, \lambda) T_2(i, j, \lambda) Q E(i, j, \lambda) d\lambda \quad (51)
\]

The spatial dependence of the optics and filter transmission coefficients and that of QE can be lumped together into one quantity and taken out of the integral sign. This quantity is called flatfield, \( FF(i, j, f_1, f_2) \), and is dependent on the filter combination used. It is normalized according to the equation

\[
\frac{1}{N^2} \sum_{j=1}^{N} \sum_{i=1}^{N} FF(i, j, f_1, f_2) = 1.0
\]

The resulting equation of \( e_p(i, j) \) can then be divided by the gain value, \( g \), used to get the corresponding DN value of the detected signal. The optics collecting area and solid angle are known quantities, however, the transmission coefficients, exposure time sample dependency, quantum efficiency, and gain are all measured in the lab. Thus, a correction factor is introduced to account for errors in any measurements taken during ground calibration as well as changes in sensitivity over time. This correction factor is determined from in-flight measurements taken of standard stars and other bodies whose intensity has been independently determined. With these adjustments, eq. (51) becomes

\[
DN(i, j) = \frac{A \Omega t(i) FF(i, j, f_1, f_2)}{g} \int I(i, j, \lambda) T_o(\lambda) T_1(\lambda) T_2(\lambda) Q E(\lambda) d\lambda \quad (52)
\]
The above equation for measured DN considers only an ideal case. In reality, there are other processes that can affect the signal level recorded in a pixel. These are discussed below.

CCD chips produce a positive DN value at zero exposure, so bias subtraction is one of the basic CCD image processing steps. In addition to this, ISS images have a coherent noise that is noticeable at low signal levels and appears as a horizontal banding pattern. The exact origin of this “2Hz-noise” is not known but seems to arise from the A/D converter during readout. The power spectrum of the noise has two main peaks in the 2-3Hz range in NAC images and one main peak near 4Hz in WAC images. The noise varies by a few DNs in gain state 2 while it only changes by about 1 DN and 0.5 DN in gain states 1 and 0, respectively. Moreover, in binned images, the noise takes the shape of a more complicated diagonal pattern. To account for the bias, a line dependent term, \( D\text{N}_0(j) \), should be added to the above equation.

Dark current in ISS does not result from the traditional signal that arises from thermal electrons. At a detector operating temperature of \(-90^\circ\text{C}\), the contribution of thermal electrons to the signal is minimal. Instead, a large part of the dark signal in the ISS comes from Residual Bulk Image (RBI). This effect arises from build-up of electrons into traps in the silicon chip, which later leak into potential wells of subsequent images. To control this effect and ensure the same starting condition for each image, the CCDs are ringed by infrared light sources that illuminate the detector to \(\sim\) 50 times the saturation level. This “pre-flash” happens before each exposure to ensure that the traps are filled. The pre-flash is then followed by readout of the electrons in the potential well. The dark current in the detectors is, thus, mainly RBI from the pre-flash stage. As in the bias case, a separate term accounting for dark current, \( D\text{N}_D(i, j) \), must be added to eq. \((52)\).

The deviation of the detector from linearity is another aspect that must be considered. Linearity describes how closely the CCD response follows a linear relation to the incoming
light. For the ISS, non-linearity is significant only for the lowest gain state (gain index 0); it is less than 1% over most of the dynamic range of the other gain states. A more detailed discussion of the linear response of the ISS detectors can be found in section 5.1.11 of the Ground Calibration Report.

The A/D converter also introduces an error in the recorded signal due to uneven bit weighting. In this digitization error, the A/D converter favors some DN values over others, deviating from the linear increase in output DN expected as the collected charge increases. Correction for this error can only be done to 12-bit images; “LUT and lossy images cannot be corrected due to information lost in the encoding and compression processes respectively” (Knowles, 2018). Description of the NAC and WAC output DN characterization as a function of input signal can be found in section 5.1.9 of the Ground Calibration Report.

Another noise source in the detectors comes from their capability to operate with an anti-blooming mode. This mode is used to control the bleeding that can occur when a pixel is oversaturated, preventing other pixels from being affected. When anti-blooming is set to ON, excess electrons are moved into traps at the expense of the adjacent pixel down the line. This results in a series of bright and dark pixel pairs wherever saturation occurs.

Calibration of ISS images should, thus, correct the aforementioned effects and relate DN to intensity. This is done with the following steps:

1. If data was converted to 8-bit using an LUT, convert back to 12-bit using reverse operation

2. Correct 12-bit images for uneven bit weighing

3. Subtract bias using average of overclock pixels

4. Subtract 2Hz noise; there are different approaches to accomplishing this depending on the image scene and camera settings (see section 4.8)
5. Subtract dark current and RBI taking into account different camera parameters

6. Correct for bright/dark pixel pairs in images taken with anti-blooming setting turned on

7. Perform nonlinearity correction for the appropriate gain state

8. Correct for pixel-to-pixel variation in CCD response by dividing by $F F(i, j, f_1, f_2)$

9. Multiply by gain value $g$ to convert DN to electrons $e_p(i, j)$. Then divide by $A \Omega t(i)$ to get average intensity over the passband of the camera filter

10. Dividing by weighting function $\int T_o(\lambda)T_1(\lambda)T_2(\lambda)QE(\lambda)d\lambda$ will give intensity in units of (Photons cm$^{-2}$ s$^{-1}$ ster$^{-1}$ nm$^{-1}$)

11. Divide by appropriate correction factor $C(f_1, f_2)$

12. Optionally, geometric correction can be done as an added step although the geometric distortion of the cameras is minimal

The resulting intensity can be normalized by $F$, where $\pi F$ is the incident solar flux and is given by

$$F = \frac{\int F_1(\lambda)T_o(\lambda)T_1(\lambda)T_2(\lambda)QE(\lambda)d\lambda}{\pi R^2 \int T_o(\lambda)T_1(\lambda)T_2(\lambda)QE(\lambda)d\lambda}$$

(53)

where $F_1$ is the solar flux at 1 AU, $R$ is the distance from the Sun to target body in units of AU. The calibration software used in this study follows the above steps and is discussed below.

### 4.8 CISSCAL

Cassini ISS images are calibrated using the CISSCAL (Cassini ISS CALibration) software. CISSCAL was developed by the Cassini Imaging Team and is used for performing basic
calibration steps, such as bias and dark current subtraction, as well as ISS specific calibra-
tion steps like uneven bit weight correction. It runs in the IDL (Interactive Data Language)
environment and requires IDL version 5.5 or later to be installed. CISSCAL follows the
calibration steps as outlined above. It uses information in the image header to determine
if a certain calibration step should be executed or not. In addition, some steps have user-
definable parameters and/or user-determinable method of execution. The calibration steps
are presented here in the same order of implementation as in the software.

LUT conversion

Images that have been encoded to 8-bit are converted back to 12-bits using a reverse LUT
to that shown in Fig. [12]. This reverse LUT is hardcoded into CISSCAL. This calibration
step has no user-definable parameters.

Bit-weight correction

As mentioned in section [4.7] bit-weight correction is not applied on images that have been
LUT encoded or lossy compressed. Uneven bit weighting essentially results from differ-
ences in bin widths. It shows up as a periodic series of spikes in the histogram of an image
with a broad range of DN values. The uneven DN distribution is modeled and corrected
mathematically, the details of which are described in section 5.1.9 of the Ground Calibra-
tion Report.

Subtract Bias/2Hz noise

CISSCAL does bias and 2Hz noise subtraction in one step. No model has been derived that
accurately reproduces the 2Hz noise, so removal of this noise is done on an image-by-image
basis. CISSCAL has three methods of bias/2Hz subtraction that the user can choose from.
The first is the “Constant Bias” method, which subtracts a constant bias offset value from the image. During image readout in the ISS, two overclock pixels are generated before a row of 1024 pixels is read from the output register and six overclocks are generated after. These are stored as two separate sums—sum of the first two overclocks and sum of the six overclocks. Averaging of these sums for each line gives a 1D array of overclock pixels. The Constant Bias method simply subtracts the average of the overclock pixels array from the image. The 2Hz noise in LUT-encoded images is at a low enough level that quantization of the encoding makes accurate reconstruction of the noise difficult. Hence, a constant bias subtraction should be used for these images.

The second method is the “Overclocked Pixels” method. It accounts for a 2Hz signal that takes the shape of a horizontal banding pattern by subtracting a line dependent value from the image. In this method, the bias is first removed by subtracting a linear fit to the overclock pixel array, i.e. the bias offset is predicted for each row of pixels from the best-fit line to the DN vs. Line Number graph and subtracted from the image. Next the overclock pixel array is smoothed with a low-pass filter to remove high-frequency noise before applying a high-pass filter to remove the bias offset and any low frequency data. This isolates the 2Hz noise component, which can then be subtracted from the image line-by-line.

The third option of bias/2Hz subtraction is the “Image Mean” algorithm, which operates on the same principle as the Overclocked Pixels method. It only applies for images that contain dark sky across their entire vertical extent. To use this method, bright pixels must first be removed. The user can use the “Auto Mask” feature to mask bright areas of the image or supply a mask file created externally. The Auto Mask feature has two parameters which need to be supplied by the user: the “Threshold” parameter determines the DN value below which pixels are considered dark sky while the “Pixel Range” parameter determines how much smoothing is done on the image before masking. Once a masked file is cre-
ated/supplied, 2Hz removal algorithm replaces the masked pixels with the average values of pixels in the same line, applies a smoothing filter then takes the median across a line. The result is then filtered with a low pass filter and a high pass filter as in the Overclocked Pixels method before subtracting it from the image.

Determining which method to use for bias/2Hz noise subtraction is a function of camera settings (particularly summation mode, data conversion, and gain) and image coverage. The constant Bias method does not account for the 2Hz noise, thus, is only suitable for images with low 2Hz signal, such as $2 \times 2$ or $4 \times 4$ summed images. Furthermore, since the banding pattern for these summation modes takes a diagonal form, Overclocked Pixels and Image Mean options are not useful in reproducing the noise. The 2Hz signal for LUT-encoded images is at a low enough level that quantization of the encoding makes accurate reproduction of the noise difficult. Hence, for these images as well, constant bias subtraction is usually the best method. For un-summed, 12-bit or LS8B images, the Overclocked Pixels or Image Mean method can be used depending on the image scene. If the image contains continuous patches of dark sky spanning the entire vertical extent, the Image Mean option may be used. Otherwise, the Overclocked Pixels method should be used.

Subtract Dark

As described in section 4.7, dark current in the ISS cameras is mainly made up of RBI from pre-flash. Leakage of electrons into potential wells happens on a comparable time scale to image readout, so the longer a charge sits on a chip, the more RBI it accumulates. How long a charge sits on the chip is determined by exposure time and readout rate. Readout rate is a function of several camera settings and can change in the middle of image readout. Hence, dark current is dependent on a complicated set of camera parameters, such as summation mode, compression mode, compression ratio, telemetry rate, etc. For this reason, taking dark frames in all combination of camera settings is not practical. Instead a modeling
approach is used.

Dark frames are taken at eight discrete exposure times spanning the entire exposure range of the cameras (0, 10, 32, 100, 220, 320, 460, and 1200 seconds). These frames are simply regular images taken with a closed shutter and are preceded by pre-flash and readout. RBI leakage as a function of time is derived from these images for each pixel, which can then be interpolated in the time domain. Now RBI continues to happen during readout as well. As each potential well is shifted vertically during readout, it picks up more charge while going through pixels downstream. Hence, the RBI contribution of pixel \([i, j]\) is the sum of RBI generated at that pixel plus all pixels down line \(j\)

\[
D_{i,j} = \sum_{k=1}^{j} RBI(i, k, t_2) - RBI(i, k, t_1)
\]

(54)

where \(t_1\) and \(t_2\) are the times when the potential well that originated at pixel \([i, j]\) enters and leaves the pixel location \([i, k]\), respectively (West et al., 2010). The parameter files from the dark frame measurements are included within CISSCAL.

In addition to RBI, some pixels exhibit high dark count due to defects caused by cosmic ray strikes, gamma rays, or manufacturing imperfection. These pixels, called “hot pixels,” are dynamic as new ones form while other ones disappear due to annealing of the silicon layer; however, the general trend is an increase in hot pixels over time. They are dealt with by taking dark frames at different periods in the mission and updating the dark models to include updated hot pixels. Dark subtraction is then done using the hot pixel snapshot closest in time to the image to be corrected. Because of the erratic nature of hot pixels and the limited number of opportunities available to take in-flight dark frames, the process of hot pixel removal is not perfect and may even add vertical streaks from saturated hot pixels. Therefore, CISSCAL has been equipped with the option to turn hot pixel removal off.
A-B Pixel pairs

Bright/dark (A-B) pixel pairs arise in images with anti-blooming state turned on. Therefore, this correction is applied on those images only. CISSCAL automatically identifies such pairs using the user-definable threshold parameter given in DN. If the difference between two neighboring pixel values is greater than the threshold value, the pixels pair is identified as bright/dark and replaced by an average of each pixel’s horizontal neighbors.

Linearize

Non-linearity correction factors were derived for both cameras at each gain state during ground calibration. These were derived by taking several images at various exposures and fitting the observed DN to exposure time. A weighting scheme was used in the fitting process that favored low DN values since they are expected to have the least deviation from linearity. The results of the analysis have been incorporated into CISSCAL.

Flatfield

Any pixel-to-pixel variations that are present are corrected during this step. This is done from data taken during ground and in-flight calibration. During ground calibration, the detector was exposed to a uniformly illuminated light source whose radiance was compared to the measured DN. A slope term relating this measured DN to the independently determined radiance level was obtained from a best linear fit and recorded in “slope files.” These files were produced for each filter combination. Additionally, flatfield measurements were taken during flight to account for shifts in the components that are likely to occur after launch. The main corrections from in-flight measurements were for dust ring removal that were identified during Venus and Titan flybys.

During this correction step, the flatfield algorithm of CISSCAL first reads the appro-
appropriate slope file and normalizes it to the average of the central 400×400 pixels. The input image is then divided by this normalized flatfield. Next, dust ring corrections are made if necessary.

Convert DN to flux

Calibration steps 9 and 10 listed in section 4.7 are done at this time. The user has the option to select the output to be in intensity units (photons/(cm² s ster nm)) or to normalized it by some flux quantity. This quantity can come from: 1) a user-provided input or 2) an auto-generated solar flux value at the distance of the target from the sun; the user can select “Jupiter” or “Saturn,” which causes CISSCAL to automatically calculate the solar distance to the target based on the image time found in the image header.

Absolute correction

Discrepancies between measured and expected signal from ground calibration are dealt with by applying a filter specific absolute correction factor. Such correction factors are determined from absolute calibration observations taken of photometric standard stars and Solar System targets including Jupiter, Enceladus, Rhea, and Dione. One of the important sources of uncertainty in the ground calibration measurements is QE, hence, a wavelength dependent QE correction has also been derived. This correction is applied to the ground based QE value during intensity calculation in the above step. However, recent absolute calibration analysis has revealed sensitivity decline over time on both detectors. Maximum declines of ~8% and ~3% were observed in the NAC and WAC, respectively. As of CISSCAL 3.9, an option has been included to correct this sensitivity decline. A complete description of the absolute correction analysis is included in Knowles (2018).
Geometric correction

In addition to the above required calibration steps, a distortion transformation on the image array can be done to account for geometric distortion caused by the optical system. Fortunately, geometric distortion in the detectors is small: maximum distortions in the NAC were measured to be 0.45 pixels in image corners and 3.36 pixels in the WAC. Thus, this correction can be done only in cases where high geometric accuracy is required.
5 Method

5.1 Data Selection and Reduction

Data products returned from NASA’s planetary missions are archived in the Planetary Data System (PDS). These products can be accessed electronically using different options. One such option is through PDS’s Planetary Image Atlas search tool:

http://pds-imaging.jpl.nasa.gov/search/

The PDS Image Atlas allows users to search and query an entire dataset of images using keywords related to the data, such as target coverage, lighting geometry, and product type.

Enceladus ISS images were selected for this study based on the following criteria:

- All images un-summed, losslessly compressed, 12-bit
- Pixel scale ≤ 12 km/pixel
- Sub-spacecraft phase angle ≤ 90 degrees
- Northernmost latitude ≤ 40 degrees

Table 3 summarizes the keywords and search values used to query data to be used in the study. The rationale behind the search criteria was to provide a loose set of constraints that will result in a large number of images from which further refinements will be made. The query returned 810 images, of which 282 were eliminated, since they were deemed low quality or were mostly dark sky images with the target covering only a small portion of the scene. In addition, the Hapke model discussed in section 3 ignores polarization of light, so images taken with a polarizing filter must be removed. A total of 108 such images were found and removed, leaving 420 images to be analyzed.
Table 3: Keywords and search values used for querying data to be used in the study.

<table>
<thead>
<tr>
<th>PDS Keyword</th>
<th>Operator</th>
<th>Value [units]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATLAS_MISSION_NAME</td>
<td>=</td>
<td>CASSINI</td>
</tr>
<tr>
<td>ATLAS_INSTRUMENT_NAME</td>
<td>=</td>
<td>ISS</td>
</tr>
<tr>
<td>TARGET_NAME</td>
<td>=</td>
<td>ENCELADUS</td>
</tr>
<tr>
<td>INSTRUMENT_MODE_ID</td>
<td>=</td>
<td>FULL</td>
</tr>
<tr>
<td>INST_CMPRS_TYPE</td>
<td>=</td>
<td>LOSSLESS</td>
</tr>
<tr>
<td>PIXEL_SCALE</td>
<td>≤</td>
<td>12 [km/pixel]</td>
</tr>
<tr>
<td>PHASE_ANGLE</td>
<td>≤</td>
<td>90 [deg]</td>
</tr>
<tr>
<td>NORTHERNMOST_LATITUDE</td>
<td>≤</td>
<td>40 [deg]</td>
</tr>
</tbody>
</table>

5.2 Data Processing

5.2.1 Calibration

After data selection, images were radiometrically calibrated using the CISSCAL software (version 3.9.1). All the calibration steps described in section 4.3 were performed. For 12-bit images, bias/2Hz noise correction was done using the Overclocked Pixels method whereas for LUT-encoded images the Constant Bias option was used. The intensity values were normalized by the incident solar flux, so that I/F of each pixel was obtained.

5.2.2 Data Import and SPICE

The rest of the image processing steps were done using the Integrated System for Imagers and Spectrometers (ISIS; version 3.9.0) software package. ISIS is a freely available digital image processing software developed by USGS to process images from current and past NASA missions. ISIS consists of many programs used to accomplish different tasks. The calibrated images in this study were imported to ISIS along with the uncalibrated versions. The image header information was then copied from the uncalibrated images to the calibrated ones.

Once images have been converted to ISIS format (called cubes), additional navigation
and ancillary data is required to perform geometric or photometric processing. For instance, cubes with supported camera models in ISIS require information such as spacecraft and target ephemerides and orientation, Sun position, and target shape model to calculate latitude/longitude and photometric angles of points on the body. Such ancillary data are made available by NASA’s Navigation and Ancillary Information Facility (NAIF) through its “SPICE” information system. One of the main components of SPICE (Spacecraft & Planetary ephemerides, Instrument, C-matrix and Events) are the nine data files, called kernels, containing the navigation and ancillary information. The ISIS application *spiceinit* was run on the calibrated images to add the necessary SPICE information, which was obtained from spacecraft navigation data. However, camera pointing kernels obtained from this source are only approximately correct, hence, result in an inaccurate geographic information being associated with each pixel.

To get around this problem, improved pointing kernels were obtained from L. Weller (private communication). For the purpose of making the publically available Enceladus data more usable, [Bland et al., 2018] created a global photogrammetric control network from over 600 Enceladus images and produced improved camera pointing kernels. Their network was later further refined and densified by L. Weller. It used images with (1) pixel scales ranging between 50 and 500 m per pixel; (2) phase angles less than 120°; and (3) filter pairs CL1:CL2 (651nm), CL1:GRN (569nm), CL1:IR3 (928nm), and CL1:UV3 (343nm). Of the 420 images selected for the present study, 265 of them were included in the [Bland et al., 2018] data set; these images, therefore, have improved camera pointing information. The majority (162) of these images were taken in the clear filter pairs while 30 of them were in CL1:GRN, 22 in CL1:IR3 and 51 in CL1:UV3. Only clear filter images were selected for further analysis since a relatively larger subset were found to provide a varied set of observations of the SPT needed to constrain the Hapke model parameters. There were only a small subset of each of the green, infrared and UV images that covered
the SPT, and those that did had a smaller phase angle range, so they were not used in the analyses.

Because of the importance of having observations that also sample in the forward-scattering direction, the phase angle value in the PDS data query (Table 3) was set to between 90° and 120° while keeping other keywords and search values unchanged. This returned one clear filter image covering the SPT that was also found in the Bland et al. (2018) data set. The image was included in the list of images used in the present study.

5.2.3 Photometry Retrieval

Clear filter images covering the SPT were selected and photometric and geographic bands were created for each image. These bands contain the observed I/F value, phase, emission, and incidence angles, and planetocentric latitude and longitude of each pixel. Then pixels covering the funiscular plains of the SPT were trimmed out. This was done by carefully determining the latitude and longitude range of the plains and using the ISIS program camtrim to trim pixels outside of a specified range. The resulting output is then cropped to remove any invalid pixels, edge pixels with uncharacteristically high I/F values, and pixels in the night side where noise accounts for a significant portion of the signal. Moreover, images containing truncated pixels (Fig. 13) had invalid pixels in the I/F band but valid pixels in other bands. Since both I/F and photometric angles are required for analysis, such pixels were also removed. The images cubes were then exported to an ASCII file format for the next stage of analysis. Dead pixels and pixels with incidence angles exceeding 90° were also removed at this step.

To examine if there are any differences in the photometric parameters of the funiscular plains with other neighboring subunits of the SPT, the above set of steps were repeated separately for images covering the south polar reticulated plains (SPRP) region and the Baghdad sulcus flanks. The list of clear filters images used to determine Hapke model
parameters and the SPT areas from which data were collected are given in Table 4 and Fig. 14, respectively. Images used for the funiscular plains region have the largest phase angle range (14.1° to 119.5°) while those used for the reticulated plains (14.1° to 46.3°) and the flank region (45.9° to 73.2°) have a much smaller range. Furthermore, the emission angle range for the flank region is limited (20.6° to 40.2°) compared to those of the funiscular plains (15.3° to 89.1°) and the reticulated plains (7.6° to 83.9°).

<table>
<thead>
<tr>
<th>Filename</th>
<th>Average Pixel Scale (m/pixel)</th>
<th>Average Phase Angle (deg)</th>
<th>Average Emission Angle (deg)</th>
<th>Average Incidence Angle (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1500056680_2*</td>
<td>349.23</td>
<td>46.58</td>
<td>40.34</td>
<td>73.40</td>
</tr>
<tr>
<td>N1500059679_2*</td>
<td>200.03</td>
<td>46.33</td>
<td>24.73</td>
<td>61.32</td>
</tr>
<tr>
<td>N1500060345_2†</td>
<td>167.40</td>
<td>46.07</td>
<td>33.33</td>
<td>76.59</td>
</tr>
<tr>
<td>N1500061132_1*†</td>
<td>128.78</td>
<td>45.84</td>
<td>32.46</td>
<td>74.75</td>
</tr>
<tr>
<td>N1500061253_1*†</td>
<td>123.10</td>
<td>45.87</td>
<td>46.59</td>
<td>75.90</td>
</tr>
<tr>
<td>N1500061512_1‡</td>
<td>110.02</td>
<td>46.03</td>
<td>20.64</td>
<td>47.90</td>
</tr>
<tr>
<td>N1500062131_1†</td>
<td>79.67</td>
<td>45.70</td>
<td>26.13</td>
<td>48.32</td>
</tr>
<tr>
<td>N1500062262_1*</td>
<td>73.31</td>
<td>45.32</td>
<td>21.53</td>
<td>62.08</td>
</tr>
<tr>
<td>N1597182896_2*‡</td>
<td>63.80</td>
<td>73.17</td>
<td>23.35</td>
<td>84.96</td>
</tr>
<tr>
<td>N1604167575_2*</td>
<td>76.14</td>
<td>73.23</td>
<td>20.14</td>
<td>85.08</td>
</tr>
<tr>
<td>N1604167805_2*</td>
<td>100.55</td>
<td>73.38</td>
<td>23.26</td>
<td>69.09</td>
</tr>
<tr>
<td>N1637463944_1*</td>
<td>36.31</td>
<td>35.86</td>
<td>69.27</td>
<td>81.30</td>
</tr>
<tr>
<td>N1637465040_1*</td>
<td>86.23</td>
<td>33.61</td>
<td>62.52</td>
<td>80.20</td>
</tr>
<tr>
<td>N1637466149_1*</td>
<td>137.99</td>
<td>33.87</td>
<td>58.58</td>
<td>68.25</td>
</tr>
<tr>
<td>N1652863123_1*</td>
<td>257.75</td>
<td>28.55</td>
<td>45.52</td>
<td>38.93</td>
</tr>
<tr>
<td>N1660433525_1*</td>
<td>40.68</td>
<td>28.25</td>
<td>67.93</td>
<td>82.25</td>
</tr>
<tr>
<td>N1660433671_1*</td>
<td>46.25</td>
<td>27.11</td>
<td>68.61</td>
<td>76.11</td>
</tr>
<tr>
<td>N1660433987_1*</td>
<td>58.35</td>
<td>25.24</td>
<td>59.97</td>
<td>65.76</td>
</tr>
<tr>
<td>N1669819472_1*</td>
<td>286.33</td>
<td>14.13</td>
<td>44.07</td>
<td>54.02</td>
</tr>
<tr>
<td>N1671591961_1*</td>
<td>158.10</td>
<td>14.47</td>
<td>56.45</td>
<td>58.28</td>
</tr>
<tr>
<td>N1671592162_1*†</td>
<td>165.69</td>
<td>14.34</td>
<td>49.05</td>
<td>57.29</td>
</tr>
<tr>
<td>N1671592381_1†</td>
<td>174.23</td>
<td>14.22</td>
<td>53.31</td>
<td>66.97</td>
</tr>
<tr>
<td>N1702359885_1*</td>
<td>135.85</td>
<td>17.16</td>
<td>49.67</td>
<td>61.26</td>
</tr>
<tr>
<td>N1702360022_1†</td>
<td>138.50</td>
<td>16.25</td>
<td>56.19</td>
<td>61.00</td>
</tr>
<tr>
<td>W1637463028_1*</td>
<td>123.43</td>
<td>119.13</td>
<td>52.32</td>
<td>79.41</td>
</tr>
</tbody>
</table>

Table 4: List of Cassini ISS clear filter images from which photometric information was extracted to determine Hapke model parameters. The symbols indicate images used for the funiscular plains (*), the south polar reticulated plains (†), and Baghdad sulcus flanks (‡).
Figure 14: Regions from which fitting data were sampled for the funisicular plains (yellow), south polar reticulated plains (tan), and flanks of Baghdad sulcus (red). The black star indicates location of Enceladus’ south pole.

5.3 Hapke Fitting

5.3.1 Bayesian Inversion

Non-linear models, like the Hapke model, are often difficult to solve because finding a unique solution is a challenging, if not an impossible, task. Such problems can, however, be approached under a Bayesian framework. In this framework, the inversion problem,
which constitutes the estimation of model parameters from observations, is solved based on the concept of state of information that is characterized by probability density functions (PDFs). Tarantola et al. (1982) framed the inverse problem using PDFs instead of central estimators like the mean and showed how a priori information on the data and parameters can be used with a theoretical model describing the relationship between parameters to find solutions which are a posteriori density functions. The basic concept of their formulation is described below:

- **Theoretical relationship:** One of the required components is a model, $G$, to compute simulated data, $d$, from parameters, $p$:

\[ d = G(p) \]

The theoretical relationship is represented by a density function $\theta(d, p)$. In this study, the version of the Hapke model used in the inversion problem is

\[ r(i, e, g) = \frac{w}{4\pi} \frac{\mu_{0e}}{\mu_{0e} + \mu_e} \{[1 + B_{S0}B_S(g)]p(g) + H(\mu_{0e})H(\mu_e) - 1\}S(i, e, \psi) \tag{55} \]

where $w$, $H(x)$, and $B_S(g)$ are given by eqs. (18a), (30a) and (39b), respectively; $\mu_{0e}, \mu_e$, and $S(i, e, \psi)$ are given by eq. (42) or eq. (43), appropriately. Empirical models are often used to describe particle phase functions, and the most common model used is the two-term Henyey-Greenstein function (Henyey and Greenstein, 1941)

\[ p(g) = (1 - c) \frac{1 - b^2}{(1 + 2b \cos g + b^2)^{3/2}} + c \frac{1 - b^2}{(1 - 2b \cos g + b^2)^{3/2}} \tag{56} \]

The first term describes the forward scattering lobe and the second the backward lobe.
The parameters $b$ and $c$ determine, respectively, the shape and backscattering fraction of the lobes. Both parameters are constrained to the range $[0,1]$. A forward scattering particle has $c < 0.5$ and a backscattering particle has $c > 1$; a high value of $b$ results in a high and narrow lobe while a low value gives a low and wide scattering peak.

The simulated data is $r$, and the parameters to be determined are $w, b, c, \bar{\theta}, B_{S0},$ and $h$. It is important to note that the model is assumed to be exact, i.e. errors in the theory are ignored.

- **A priori** information: The observational data and any knowledge on the parameters constitute a priori information on the system. The data is described by a PDF, $\rho_d(d)$, that is assumed here to be Gaussian, with dimension $N_g$ equal to the number of viewing geometries. This is a simplifying assumption made in the inversion process. The PDF, $\rho_d(d)$, is made up of individual observations, $d_{obs}^i$, at geometry $i$ with mean $d_{obs}$ and covariance matrix $C$ describing the Gaussian experimental uncertainties. An observation at a given geometry is assumed to be independent of observations at other geometries, so $C$ is a diagonal matrix with $C_{ii} = \sigma_i$ being the standard deviation of the observation $i$. The standard deviation was taken to be 10%, $\sigma_i = (1/10) \times d_{obs}^i$, or 0.01, whichever is greater.

The **a priori** information on the parameters, $\rho_p(p)$, is independent of that on the data and, in the present case, is taken to be null. The only prior knowledge one has on the parameters is their domain of existence: $[0,1]$ for $w, b, c, B_{S0}$, and $h_S$ and $[0,45^\circ]$ for $\bar{\theta}$. Therefore, $\rho_p(p)$ has a constant value in the intervals and is zero outside the bounds.

- The conjunction of the **a priori** density function with the theoretical relationship gives the **a posteriori** state of information, which is the solution to the inverse problem. The posterior density function on the parameters $\sigma_p(p)$ is given by the equation
\[ \sigma_p(p) = k \rho_p(p) L(p) \]  

(57a)

where \( k \) is a normalization constant; \( L(m) \) is a “likelihood function” that is a measure of the fit between the observed and modeled data and is given by

\[ L(p) = \int \frac{\rho_d(d) \theta(d|p)}{\mu_d(d)} dd \]  

(57b)

where \( \theta(d|p) \) is the theoretical PDF of the simulated data \( d \), given \( p \); since the theoretical relationship is assumed to be exact, \( \theta(d|p) = \delta(d - G(p)) \). The quantity \( \mu_d(d) \) is the PDF of the state of null information on the data, which is simply a uniform PDF in the present case.

### 5.3.2 Monte Carlo Markov Chain (MCMC)

The integration in eq. (57) can be solved analytically for linear problems. For a non-linear problem like the Hapke model however, the Monte Carlo approach using the Metropolis rule \( \text{[Metropolis et al. 1953]} \) should be employed to sample a posteriori density distribution \( \text{[Mosegaard and Tarantola 1995]} \). This approach has previously been used to estimate Hapke model parameters for the Martian surface \( \text{[Fernando et al. 2013]} \) and to study uncertainties of Hapke model parameters \( \text{[Schmidt and Fernando 2015]} \). Their MCMC software was used in this study (F. Schmidt, private communication). The MCMC method involves randomly generating parameter values within the model space and accepting proposed values with a probability given by the Metropolis rule. This results in a Markov chain which, after a sufficient number of iterations, equilibrates to the posterior distribution. The mean and standard deviation of the distribution can then be calculated to describe the value and
uncertainty of each parameter. It is worth noting that the \textit{a posteriori} probability density does not have to be Gaussian; it could be multimodal or have any other shape.

How well a given parameter has been resolved by the observed data is an important concept to consider in the results of an MCMC. If a parameter’s \textit{a posteriori} PDF is the same as its \textit{a priori} PDF, then there is no information gained from the observation in regards to the parameter in question. To test the resolution of a parameter, Fernando et al. (2013) and Schmidt and Fernando (2015) developed a non-uniformity criterion \( \hat{k} \). Central moments \( \mu_n \), like the mean and variance, are often used in central estimators but a cumulative moment function with cumulants \( k_n \) has the advantage of easily giving unbiased central estimators from a sample of a population (Fisher, 1930). For a uniform PDF, the first four cumulants \( k_1, k_2, k_3, k_4 \) are \( 1/2, 1/12, 0, \) and \( -1/120 \), respectively. Therefore for a resolved parameter, the corresponding cumulants of its posterior PDF should be different from these values. Fernando et al. (2013) and Schmidt and Fernando (2015) proposed an expression for \( \hat{k} \) given by

\[
\hat{k} = \max \left| \frac{k_1 - 1/2}{1/2}, \frac{k_2 - 1/12}{1/12}, \frac{k_3}{1/60}, \frac{k_4 + 1/120}{1/120} \right| \tag{58}
\]

They conducted a numerical test of 10,000 uniform random vectors of 500 samples and found a maximum \( \hat{k} \) value of 0.47. Hence, a lower limit of \( \hat{k} = 0.5 \) is proposed to distinguish (un)resolved parameters, i.e. parameters with \( \hat{k} < 0.5 \) for their \textit{a posteriori} PDF are considered unresolved by the observations and vice versa.
6 Fitting Results

The mean, standard deviation, and non-uniformity criterion of each of the six parameters were determined for each geological subunit. Despite the absence of data near opposition, the SHOE parameters, $B_{S0}$ and $h_S$, have been kept in all the inversions. This was done for two reasons: (1) Fernando et al. (2013) tested whether fixing these parameters to zero would have any change in the estimation of the other parameters and observed no change, and (2) the nature of the inversion algorithm means fixing $B_{S0}$ and $h_S$ does not save on computation time. However, the physical interpretations of these parameters are made with caution since data near opposition are required to accurately constrain them.

6.1 Funisicular Plains

Several parameter estimates were made for the funisicular plains region by varying the number of iterations in the MCMC from 10,000 to 200,000, with a burn-in phase (consisting of the discarded steps which the Markov Chain takes before reaching a stationary state) of 5000 iterations. The inversion results gave two sets of solutions for each parameter (Table 5) with the PDF showing, except for the $h_S$ parameter in Fig. 15b, one peak near the mean value (Fig. 15). There were also instances where the PDFs of $b, c, w$, and $\bar{\theta}$ all took-on a bimodal shape, with one peak near each of the mean values given in Table 5. In such cases, the SHOE parameters had either a bimodal or multimodal distribution. The non-uniform posterior distribution of the parameters $b, c, w$, and $\bar{\theta}$, as shown by their PDFs and $\hat{k} > 0.5$, means a solution exists for the four parameters. Although the MCMC comes to a solution for $h_S$ and $B_{S0}$ parameters, the values are not constrained.

Comparison of the retrieved parameter values to laboratory and other orbital measurements gives some indication as to which of estimates is the more believable one. McGuire and Hapke (1995) studied the light scattering properties of a variety of artificial particles...
Table 5: Representative examples of the two sets of retrieved Hapke model parameters ($\theta$ in degrees) and their standard deviations (in parenthesis) for the funisicular plains from 45,000 iterations (Estimate 1) and 95,000 iterations (Estimate 2). The value of the non-uniformity criterion, $\hat{c}$, is also given (third line of each entry). Corresponding PDF of each parameter is shown in Fig. [15].

which differed in shape, surface roughness, absorption coefficient, and density of internal scatterers. They found the two-parameter Henyey-Greenstein (2PHG) function (with parameters $b$ and $c$) adequately described the phase functions of the studied particles. Moreover, a plot of $c$ vs. $b$ revealed that the two parameters are highly correlated; it showed the physical characteristics of different particle types map into specific areas of an L-shaped region. In general, smooth clear spheres were found to be strongly forward scattering (low $c$) with narrow scattering lobes (high $b$). Departures from sphericity, due to roughness or irregularity, take part of the light from the forward lobe and redistribute it to other directions. In addition, increased density of internal scatterers leads to a diminishing of the forward lobe. These changes result in a scattering pattern with a broader (low $b$) and stronger backscattering (high $c$) lobe. Subsequent studies by other authors have used 2PHG functions to fit measured phase functions and observed the inverse relation between $b$ and $c$. Hapke (2012a) has collated and plotted $b$ and $c$ values reported in a number of studies, which come from 495 measurements of a variety of particle types including mineral separates, volcanic soils and their derivatives, meteorites, and lunar, Martian, and Europan regoliths, as well as the McGuire and Hapke (1995) artificial regolith analogs. With the
Figure 15: Corresponding PDF of each of the six parameters given in Table 5 from 45,000 (a) and 95,000 (b) iterations. Vertical lines indicate the initial values of the parameters at the beginning of the Markov Chain.
Figure 16: Plot of $c$ vs. $b$ and the empirical “hockey-stick” relation (eq. (59); solid line) for (a) the 495 data values collated by Hapke (from Hapke (2012a)) (b) the fitting results in this study showing both estimates for funiscular plains (diamond), south polar reticulated plains (asterisk), and Baghdad sulcus flanks (triangle).

exception of a few outliers, all points were found to lie along a similar L-shaped region as found by McGuire and Hapke (1995), clustered around the empirical curve

$$c_{MH} = 3.29 \exp(-17.4b^2) - 0.908$$

where $c_{MH} = 2c - 1$. The $c_{MH}$ vs. $b$ plot along with the derived empirical curve from 495 measurements has been adapted from Hapke (2012a) and is shown in Fig. 16a. A similar graph is shown in Fig. 16b with the $c_{MH} - b$ values from this study plotted. For the funiscular plains, it is clear that the values from the second estimate are outliers. In fact, out of the three groups of outliers identified by Hapke (2012a), one is a group of points with values of $b$ nearly equal to one. He notes that such values result in a phase function with such extremely high and narrow lobes (Fig. 17b) as to be unreasonable.
is also not clear how such particles might be constructed. The $b$ and $c$ values from the first estimate on the other hand result in a scattering pattern with much wider and lower peaks (Fig. 17a). Since $c > 0.5$, the backscattered peak is higher than the forward scattered peak. This is in agreement with previous studies (Buratti, 1985; Verbiscer et al., 2005; Verbiscer and Veverka, 1994), which have found Enceladus to be moderately backscattering. All of these studies lacked observations in the forward direction ($g > 90^\circ$), however, and used a single-parameter Henyey-Greenstein phase function. Therefore, an equivalent quantitative comparison cannot be made with the results from this study.

![Figure 17](image)

**Figure 17:** Volume-averaged single particle phase function with $b = 0.349$, $c = 0.555$ (a) and $b = 0.9505$, $c = 0.9995$ (b).

For the single-scattering albedo, the value from Estimate 1 is much lower than what would be expected for a bright object like Enceladus. The geometric albedo is defined as the ratio of the integral brightness of a planetary body at opposition to that of a flat disk of the same size that scatters light equally in all directions. With a geometric albedo of 1.4 (Verbiscer et al., 2007), Enceladus is the most reflective body in the Solar System. Therefore, a single-scattering albedo value close to one is expected. The value from the second estimate, thus, seems more accurate and indeed is more comparable to values reported in
previous studies. Using HST observations made with F675W filter (\(\lambda_{\text{eff}} = 672\) nm), covering sub-observer points between \(-4^\circ\) to \(-29^\circ\) latitude and phase angle range of \(0.24^\circ\) to \(6.4^\circ\), Verbiscer et al. (2005) found \(w = 0.995 \pm 0.001\). Verbiscer and Veverka (1994) also found a similarly high value of \(0.998 \pm 0.001\) using Voyager clear filter observations (480 nm) supplemented with Earth-based CCD observations made with V (550 nm) and R (634 nm) filters; the sub-Earth/spacecraft points for these observations were about \(+30^\circ\) latitude and had a phase angle range of \(0.65^\circ\) to \(43^\circ\). It is important to note that Verbiscer and Veverka (1994) used a modified Hapke model to that used here that included exact solutions to anisotropic multiple scattering. Verbiscer et al. (2005) also used this exact solution but on Hapke’s 2002 model, which incorporates opposition effects due to coherent backscattering. The macroscopic roughness values from both estimates characterize a rough surface, which is consistent with the ropy texture of the funiscular ridges. This is in contrast to the much lower values reported by Verbiscer and Veverka (1994) (\(\bar{\theta} = 6^\circ \pm 1^\circ\)). Verbiscer et al. (2005) also found a similarly low value (\(\bar{\theta} = 7^\circ \pm 3^\circ\)) from data collected using HST F439W filter (\(\lambda_{\text{eff}} = 434\) nm) and Voyager clear filter, which together covered phase angle range of \(0.29^\circ\) to \(43.5^\circ\). Hence for the macroscopic roughness, the maximum phase angle observations in both of these studies were made at about \(43^\circ\). However, according to Buratti and Veverka (1985), the effects of topographic roughness become significant starting at phase angles of 30-40°, and Helfenstein (1988) notes that reliable determination of photometric roughness requires observations which extend from small phase angles out to phase angles greater than \(90^\circ\). Therefore, the lower roughness values reported are likely to be not well constrained.

As would be expected from the lack of low phase angle observations in this study, the SHOE parameters are poorly constrained. In both estimates, the value of \(h_S\) is near the set upper limit of 1. Recalling that \(h_S\) represents the angular-width of the SHOE peak in radians, and the HWHM occurs at \(2h_S\), the upper limit corresponds to a HWHM that
is unrealistically large (115°). In reality, it is unlikely that any detectable SHOE HWHM would be greater than 90°, so \( h_S \) values should be less than about 0.8 (Verbiscer et al., 2018). Likewise, the SHOE amplitude parameter is inconsistent with the average particle phase function parameters. Since \( B_{S0} \approx S(0)/\wp(0) \), for an opaque particle all the light comes from the illuminated portion, so \( B_{S0} = 1 \). Hence, for such a particle, a forward scattered component would not be expected. For Estimate 1, this is in contradiction with the predicted scattering pattern from the particle phase function parameters (Fig. 17a), which has a forward scattered component. Estimate 2 predicts regolith grains that are highly transparent, but the result is reported with caution. For comparison, Verbiscer et al. (2005) found Enceladus’ surface to be nearly opaque, with \( B_{S0} = 0.98 \) at \( \lambda_{eff} = 672 \) nm; however, this should not be taken as a validation of the Estimate 1 value given they are using a different version of the model and their small phase angle range.

### 6.2 SPRP and Baghdad Sulcus

The estimated parameters for the SPRP and Baghdad sulcus flanks and the corresponding PDFs from the Markov Chain are given in Table 6 and Fig. 18. The estimates were made from the last 145,000 and 95,000 iterations (after a burn-in phase of 5,000 iterations) of the Chain for the SPRP and the tiger stripe flank, respectively. As in the previous case, the PDFs and \( \hat{k} \) values show a non-uniform distribution, so the posterior PDFs are different from the uniform \( a \text{ priori} \) distribution. Unlike the case of the funiscular plains, these estimates were made from data with a limited phase angle coverage. The lack of data in the forward scattering direction limits the ability to model the phase function parameters as well as the macroscopic roughness, so refinements of these terms will be necessary. But for now, the values are compared to those of the funiscular plains’ and previously published estimates.
The particle phase function parameters indicate the SPRP is forward scattering, with a low amplitude lobe in the backward direction while Baghdad sulcus flanks are strongly backscattering, more so than the funiscular plains. The lower \( c \) value of the SPRP suggests the grains have a lower density of internal scatterers than those of the other two regions. However, the higher \( b \) value of both the SPRP and the flank region relative to the first estimate of the funiscular plains indicates a lower degree of particle shape irregularities and particle roughness. A plot of the parameter values in the \( c_{MH} \) vs. \( b \) graph (Fig. 16b) shows, however, that they are removed from the empirical L-shaped region. This is most likely due to the lack of data at \( g > 90^\circ \). Indeed, Verbiscer et al. (2018) mention an occasional pitfall of the 2PHG function that occurs when observations are restricted to the backscattering direction with few, or in this case no, sampling in the forward direction. Because the 2PHG function assumes the forward and backward scattering lobes have equal widths, little to no sampling in the forward direction results in a forward-scattered lobe that is being controlled by the fitted shape of the backscattering lobe.

The single-scattering albedo of the SPRP is typical of a bright surface like Enceladus and is in good agreement with the results of Verbiser et al. (2005) and Verbiser and
Figure 18: Corresponding PDF of each of the six parameters given in Table 6 for (a) SPRP and (b) Baghdad sulcus flanks. Vertical lines indicate the initial values of the parameters at the beginning of the Markov Chain.
Only Neptune’s satellite Triton, with an average global \( w = 0.998 \) (Hillier et al., 1994), has a similarly high single-scattering albedo. The flanks of Baghdad sulcus on the other hand have a much lower single-scattering albedo. This result is consistent with previous ISS NAC observations, which show a dark material extending a few kilometers to either side of the tiger stripes (Porco et al., 2006).

The SPRP macroscopic roughness was found to be \( \theta = 4.0 \pm 0.1 \), in close agreement with the values found by Verbiscer et al. (2005) and Verbiscer and Veverka (1994). However, similar to these studies, the phase angle coverage of the SPRP spans to about 46° so the value found here may not be well constrained. Although the measurement lacks data at \( g > 90° \), the flank region is found to have a higher roughness, with \( \theta = 21.1 \pm 0.4 \), but not as high as that of the funiscular plains. This is in contrast to the results of Annex et al. (2012), who found the tiger stripe flanks to have a higher roughness value than that of funiscular plains. Their result is consistent with multi-wavelength ISS and VIMS observations that show stronger water-ice absorption of the spectra at the tiger stripes and weaker absorption away from the fractures. The absorption band depths are compared with the spectral signature of water-ice of varying particle diameters and indicate largest particle sizes at the “fresh” surface material of the tiger stripes and smaller sizes towards the relatively older material away from the tiger stripes (Jaumann et al., 2008; Porco et al., 2006). Given the positive correlation between the macroscopic roughness parameter and grain size that has been shown in laboratory study of planetary analogues (Cord et al., 2003), the coarse grained tiger stripe would be expected to have a higher roughness. The fact that this is not the case in the present study could be due to the limited range of observational geometry, particularly for the Baghdad sulcus flanks, or some inherent surface property. Verbiscer et al. (2018) note that the estimation of the macroscopic roughness parameter using a disk-resolved model requires a good range in emission angles for reliable fits. Unfortunately, the SPT is particularly difficult to model mainly because of the limited range in emission
angle.

The angular width of the SHOE parameter for both the SPRP and the tiger stripe flank region is similarly high as the Table 5 values for the funicular plains and is thus unrealistic. The low value for the amplitude parameter of the SPRP suggests highly transparent grains, in agreement with the low density of internal scatterers indicated by $c$ parameter, while the high value of the Baghdad sulcus flanks suggests opaque surface particles, also in agreement with the high density of internal scatterers indicated by the $c$ parameter and the backscattering nature of its phase function. However, it must be stressed that low phase angle observations near opposition and high phase angle observations in the forward scattering direction are needed to better constrain the SHOE and phase angle parameters, respectively.
7 Conclusion

Disk-resolved Cassini ISS clear filter images of the SPT of Enceladus have been used to determine photometric parameters using Hapke’s (1993) equation and the improved approximation to the Ambartsumian-Chandrasekhar $H$-function (Hapke, 2002). Three sub-units of the SPT, the funiscular plains, south polar reticulated plains, and Baghdad sulcus flanks, were separately fit to the model under a Bayesian inversion framework using data spanning phase angle ranges of 14° to 119°, 14° to 46°, and 46° to 73°, respectively. The lack of data near zero phase means the SHOE parameters cannot be accurately constrained from the data set used in this study. Despite having the largest phase angle coverage, the results of the other four parameters for the funiscular plains were not clear-cut, with various iterations of the inversion process showing peaks at two sets of values in the posterior PDF. Comparison with laboratory data and spacecraft/telescopic observations suggests that the phase function parameters are better described by Estimate 1 values while the single-scattering albedo is better described by the Estimate 2 value. While both estimates of the roughness parameter are within the physically realistic range and describe a rough surface that is consistent with the ropy texture of the funiscular ridges, additional sampling, particularly in the forward scattering direction, is required to better constrain the parameter.

Inversion results for the SPRP and Baghdad sulcus flanks showed consistent peaks at different iterations. The phase functions show particles in the tiger stripe flanks are fully backscattering while those in the SPRP are forward scattering with a small backscattering component. This suggests a high density of internal scatterers in particles near the tiger stripe fractures and a low density in SPRP particles. The funiscular plains, with a moderately backscattering phase function, have a medium density of internal scatterers, but its low $b$ value means it has a higher particle shape irregularity. The single-scattering albedo values indicate the SPRP is the brightest of the three terrain types while the tiger stripe
flanks are the lowest. On the macroscopic scale, the SPRP, which is farthest from the tiger stripes and thus receives less fresh deposits from the active vents, has the lowest roughness value while the funiscular plains and the tiger stripe flanks are much rougher. This can be explained by the exposure of the older surface deposits at the SPRP to surface weathering processes, such as sputtering and comminution due to high-velocity impact of E-ring particles, for a longer period of time leading to smaller particle sizes, and thus roughness value. Larger particles emanating from the active jet sources tend to be concentrated close to the vents as they are heavy; although this newly deposited material is also subjected to the same weathering processes, their shorter exposure time means they are relatively coarser than older deposits away from the fractures. Comparison of the funiscular plains and Baghdad sulcus roughness values, however, contradicts this; if particle size is a proxy for macroscopic roughness, then the tiger stripe region should have a higher roughness. This is most likely due to the limited range of phase and emission angles in the data used to fit the Baghdad sulcus region.

The most important limitation of this study is the lack of data near opposition and the scarce sampling in the forward direction. Low phase angle observations are required to constrain the opposition effect parameters, while high phase angle observations constrain the directional scattering and macroscopic roughness parameters. A more complete observational geometry would, thus, be required to resolve some of the inconclusive, and sometimes inconsistent, results of this study. Attaining a good fidelity of parameter values is not only important for making accurate physical interpretations of Enceladus’ surface but also extrapolating the photometric behavior to wavelengths outside the observational range. This would be essential, for instance, for thermal modeling of the surface.

The placement of the SPT at the south pole of Enceladus makes obtaining a diverse set of observational geometry a challenge. However, it is still possible to maximize the use of the data that has already been collected by the ISS if improved camera pointing
kernels are made available. This effort has already been started by Bland et al. (2018), who provided updated pointing kernels for images from a global photogrammetric control network. Because of the SPT’s scientific importance, however, future control network development for Enceladus should be directed towards this region.
References


