PHYSICALLY BASED LIGHTING CALCULATIONS FOR COMPUTER GRAPHICS

BY

PETER S. SHIRLEY

B.A., Reed College, 1985

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1991

Urbana, Illinois

©Copyright by

Peter S. Shirley

1990

ABSTRACT

Realistic image generation is presented in a theoretical formulation that builds from previous work on the rendering equation. Previous and new solution techniques for the global illumination are discussed in the context of this formulation.

The basic physics of reflection and light transport are used to derive the rendering equation. The problem of generating an image is phrased in terms of evaluating the Global Radiance Function, which consists of radiance values for all points and directions. This formulation of the rendering equation differs from previous formulations by rigorously accounting for transparent surfaces.

The physical rules governing reflection are used to make improvements in reflection models. In diffuse transmission it is shown that light is filtered to the same extent regardless of which side of the surface the light comes from. This eliminates one of the parameters from previous diffuse transmission models. The microscopic structure of polished surfaces was used to justify coupling the diffuse and specular coefficients according to the Fresnel Equations. The Fresnel Equations are commonly used to vary the reflectivity of metal and transparent dielectrics, but have not been used before to vary the reflectivity of the polish and underlying diffuse substrate.

Image-based solution methods are phrased as a lazy evaluation of the Global Radiance Function; evaluation takes place for visible points. Several constraints were outlined for what part of the image function should contribute to each pixel, and a separable, symmetric filter is developed that satisfies these constraints.

A stochastic shadow ray generation method is introduced that reduces the number of shadow rays needed for scenes with multiple light sources. The sampling distributions used for shadow rays and other dimensions of the integral are evaluated by introducing to computer graphics the notion of discrepancy from numerical integration theory. The use of discrepancy provided some insight not given by the signal processing theory traditionally used in computer graphics. As part of this discussion a new sampling scheme, N-rooks sampling, is introduced. N-rooks sampling is shown to be as efficient to generate as jittered sampling, while often outperforming Poisson disk sampling. It also can generate distributions for any positive integer number of samples, including primes.

The peculiarities of the sampling spaces used in distributed ray tracing are shown to preclude naive hierarchical sampling. It is demonstrated that hierarchical sampling can greatly reduce noise, however, if we have sufficient knowledge of the sampling space.

Zonal methods represent the opposite extreme of image methods, where all function values are computed and stored, and each evaluation is a table lookup. The zonal method is phrased as a transport simulation, similar to progressive refinement radiosity methods. Using this direct simulation model, it is straightforward to generate zonal methods for anisotropic reflection. This requires storing accumulated power in a directional table for each zone. A proof is given that, subject to certain constraints, only $O(N)$ rays are required for a zonal solution with N zones.

Simulation allows for surfaces which are not zoned to interact with those that are. This is a generalization of the diffuse and specular ray tracing transport work of Malley. This technique can be useful for highly complex or difficult to zone surfaces such as a human face.

The zonal solution methods can be applied to participating media in a fairly natural manner. This zonal method has the benefit of not requiring as much computation time when the medium

is sparse. This also applies to media with anisotropic scattering characteristics, but such a solution requires a large amount of storage.

Wavelength dependent solutions introduce some complications, but can be handled by traditional point sampling techniques. Time dependent solutions are easily handled by image-based solution methods, but are very difficult to apply using zonal methods.

ACKNOWLEDGEMENTS

Thanks to the members of my committee, George Francis, Dennis Gannon, Donald Hearn, and Robert Skeel. To my advisor, William Kubitz, for always being supportive and reading things on even the shortest notice. Thanks to the many people at Illinois for advice and support: Polly Baker, Jenny Barrett, Randall Bramley, Bill Brown, Jean Buckley, Gayanne Carpenter, Herbert Edelsbrunner, Parris Egbert, Brian Foote, George Frank, David Griffith, John Halton, Teresa Hughey, Ray Idaszak, Atul Jain, Ralph Johnson, Henry Neeman, Greg Rogers, Dave Semeraro, Kelvin Sung, Allan Tuchman, Peter Williams. Thanks to the many people in the graphics community who gave me information on their work, and feedback on my own: John Airey, Jim Arvo, Mark Lee, Tom Malley, Claude Puech, Holly Rushmeier, David Salzman, Stephen Spencer, Sam Uselton, and John Wallace.

TABLE OF CONTENTS

1 INTRODUCTION 1

2 BASIC OPTICS 8
- 2.1 Physical and Geometrical Optics 9
- 2.2 Smooth Surfaces 11
  - 2.2.1 Snell's Law 12
  - 2.2.2 Fresnel Equations 13
  - 2.2.3 Absorption of Transmitted Light 18
- 2.3 Rough Surfaces 19
- 2.4 Composite Surfaces 20
- 2.5 Translucent Surfaces 23
- 2.6 Participating Media 24
- 2.7 Summary 25

3 GLOBAL ILLUMINATION PROBLEM FORMULATION 27
- 3.1 Terms and Units 28
  - 3.1.1 Reflectance Terms and Units 30
- 3.2 Image and Global Radiance Functions 32
- 3.3 Rendering Equation 38
- 3.4 Summary 40

4 SURFACE REFLECTION MODELS 42
- 4.1 Diffuse Reflection and Transmission 43
  - 4.1.1 Diffuse Transmission 44
- 4.2 Specular Reflection and Refraction 45
- 4.3 Combined Diffuse and Specular Reflection 47
- 4.4 Glossy Reflection 48
- 4.5 Standard Computer Graphics Reflection Models 49
- 4.6 Summary 50

5 IMAGE SPACE SOLUTION METHODS 52
- 5.1 The Image Function 53
  - 5.1.1 Finite Aperture Camera Model 58
- 5.2 Direct Lighting 60
  - 5.2.1 Ambient Light 61
- 5.3 Whitted-Style Ray Tracing 62

5.4 Stochastic Ray Tracing . . . 65
5.4.1 Cook et al.'s Distributed Ray Tracing . . . 65
5.4.2 Kajiya's Path Tracing . . . 73
5.4.3 Shadow Ray Optimization . . . 77
5.5 Strategies for Non-adaptive Sampling . . . 79
5.5.1 Random Sampling . . . 80
5.5.2 Regular Sampling . . . 81
5.5.3 Jittered Sampling . . . 82
5.5.4 Poisson Disk Sampling . . . 85
5.5.5 N-rooks Sampling . . . 87
5.5.6 Empirical Behavior of Sampling Strategies . . . 89
5.5.7 Theoretical Behavior of Sampling Strategies . . . 91
5.6 Strategies for Adaptive Sampling . . . 97
5.7 Summary . . . 101

6 ZONAL SOLUTION METHODS . . . 103
6.1 Zonal Methods for Diffuse Environments . . . 103
6.1.1 Diffuse Zonal Methods as Linear Algebraic Equations . . . 104
6.1.2 Diffuse Zonal Methods as Light Transport Simulation . . . 105
6.1.3 Form Factor Calculation . . . 106
6.2 Optimizations for Diffuse Zonal Methods . . . 110
6.2.1 Patch and Element Substructuring . . . 111
6.2.2 Speeding Up the Emission of Power . . . 113
6.2.3 Optimal Time Complexity . . . 114
6.2.4 Proof of Neccesary Ray Density . . . 115
6.2.5 Gauss-Seidel Revisited . . . 119
6.2.6 Division into Subsystems . . . 120
6.2.7 Calculation of Direct Lighting in Viewing Phase . . . 120
6.2.8 One Level Path Tracing . . . 121
6.3 Zonal Methods for General Reflection Properties . . . 121
6.3.1 Including Specular Reflection . . . 122
6.3.2 Including General Reflection Functions . . . 123
6.4 Summary . . . 126

7 EXTENSIONS TO BASIC SOLUTIONS . . . 127
7.1 Including Participating Media . . . 127
7.2 Wavelength Dependencies . . . 129
7.3 Time Dependencies . . . 131
7.4 Summary . . . 132

8 IMPLEMENTATION . . . 134
8.1 Toolbox Classes . . . 135
8.2 Ray Tracing Primitives . . . 140
8.3 Ray-Material Interaction . . . 141
8.4 Zonal Calculations . . . 144
8.5 Viewing Calculations . . . 147
8.6 Summary . . . 148

9 CONCLUSION 150

APPENDIX 154

A RAY-SURFACE INTERSECTION CALCULATIONS 154
A.1 Spheres 154
A.2 Triangles 155

B MONTE CARLO METHODS 157
B.1 Introduction 157
B.2 Background 157
B.3 Monte Carlo Integration 159
B.4 Monte Carlo Solution to an Integral Equation 162
B.5 Monte Carlo Estimates of Sums and Linear Equations 163
B.6 Generating Random Numbers With Non-Uniform Densities 164

BIBLIOGRAPHY 167

VITA 176

CHAPTER 1

INTRODUCTION

This document is about creating realistic looking pictures with computer graphics programs. What is meant by a realistic image? Hall and Greenberg[44] write:

> *Our goal in realistic image synthesis is to generate an image that evokes from the visual system a response indistinguishable from that evoked by the actual environment.*

This is a lofty goal. Attempting to give the viewer *exactly* the same metal impression from a synthetic image as that from viewing the real scene would probably require connecting into the wiring between the eyes and the brain of the viewer. This is not currently feasible.

For practical reasons, we must aim considerably lower than completely duplicating the perceptual response. The primary limitations are that computer generated images are usually displayed in only two dimensions, with limited visual field resolution and color palette, such as on a CRT or printed paper. These limitations also hamper realist painters and photographers. Large field-of-view displays used for flight simulation, and head-mounted displays decrease the limited visual field problem, but their cost and complexity has precluded common use.

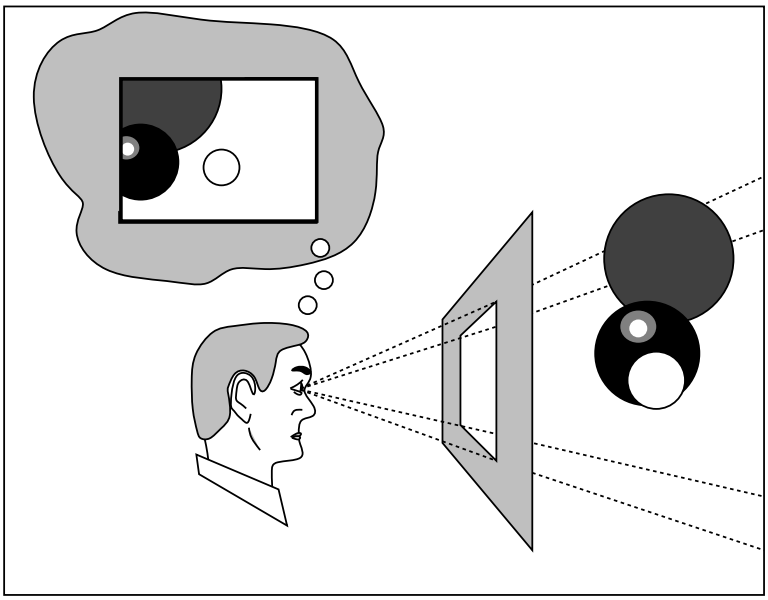

Figure 1.1: A simplified model of mental image processing.

There are two simple ways that these issues are circumvented in computer graphics. The first is shown in Figure 1.1, where a blockade having a rectangular hole is suspended directly in front of one eye of the viewer (the other eye is assumed to be closed). The compromise here is to redefine the objective from duplicating the perceptual response to duplicating the light intensity coming through the rectangular 'window'. The light intensities coming through the hole can be calculated and displayed directly on a CRT. The fundamental problem with this window method is that peripheral vision is ignored; the shift from a continuous full field-of-view 3-D image to a rectangular 2-D image framed by whatever happens to surround a CRT in a terminal room could profoundly change the visual impression of the scene. This simplified viewer model is *not* the same as the full model shown in Figure 1.2. The differences between the simplified and full model cause serious problems in many applications where aesthetic judgements are necessary.

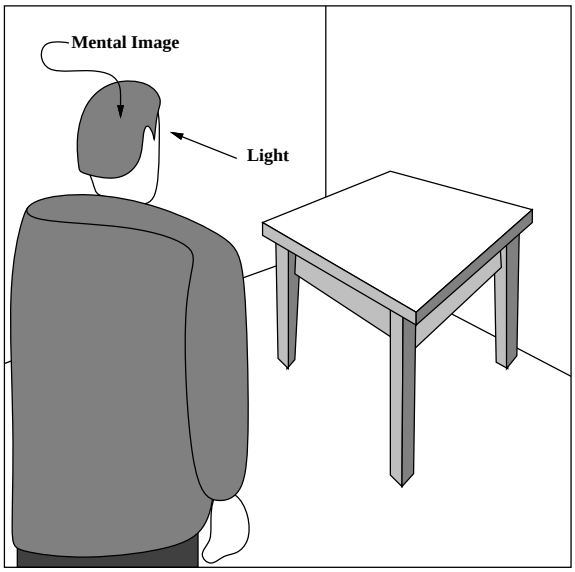

Figure 1.2: A viewer forms an image of a scene.

The other common way to circumvent the rectangular display problem is to eliminate the concept of a viewer by switching to a camera model. If we can calculate the intensities of light that would hit a piece of film in a camera then, in principle, we can determine how a photograph of the scene would appear. Because photographs are usually rectangular, we can display a 'digitized' (i.e. discretized) version of the photograph on a CRT. This greatly lessens the peripheral vision issues. They do not go away completely because the surroundings of a CRT will still affect the impressions conveyed by the picture, just as using black or white pages in a photo album can make a difference to the overall perception of the photos in the album. Figure 1.3 shows a pin-hole camera model of image acquisition. The virtual window shown in front of the camera illustrates why it makes little difference whether we are using a window model or a pin-hole camera model. The only major difference seems to be whether our image needs to be spatially inverted before display. Important differences do arise, however, when a finite aperture, which will introduce focusing effects, is used instead of a pin-hole.

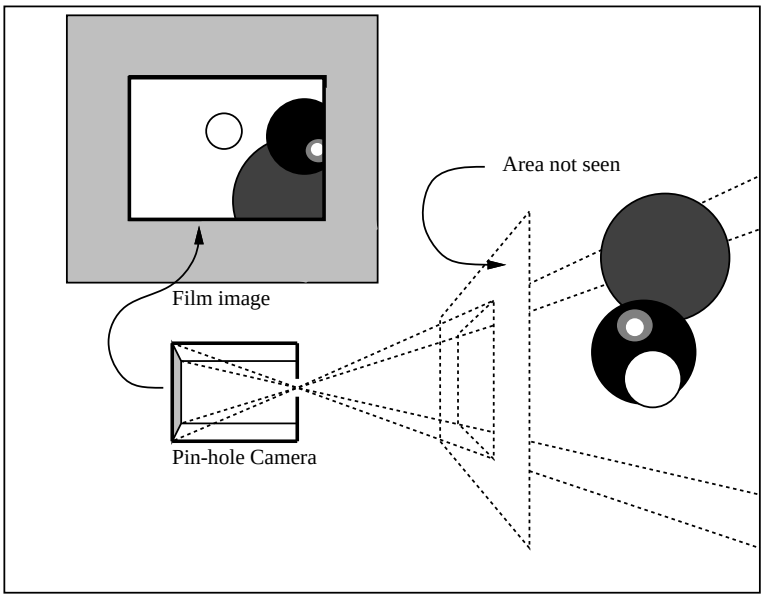

Figure 1.3: A pin-hole camera model of image acquisition.

It could be said that using the camera model is just redefining the problem; the general viewer model is simply too complicated. Thus, instead of creating an approximate impression of looking at a scene, we will create an accurate impression of looking at an approximate photograph of the scene. In spite of this, the camera model will be used for the remainder of this work. This means that the image acquisition problem is reduced to finding what light hits the film for a camera position in a particular scene. The distribution of light hitting the film can be described as the the *image function*, $I(x, y, \lambda)$. The fundamental problem attacked in this document is finding I for all points (x, y) on the film plane, and for all light wavelengths λ in the visible spectrum[1].

It should be emphasized that we have compromised considerably in switching to a camera model from Hall and Greenberg's objective of providing accurate perceptual response. A pho-

[1] If the dye of the film has response curves that include non-visible wavelengths, then we will need to include calculations for those wavelengths. Usually we assume highly idealized models of how the film operates and this issue is ignored.

tograph does generally provide an accurate portrayal of the perceptual impression of a scene, which is the reason photographers are so careful with additional artificial lighting when photographing a scene. Factoring in the characteristics of the human visual system is probably crucial in producing a *realistic* image. Although such perceptual issues are very interesting, and to some extent unexplored in the graphics literature, they are outside the scope of this work. The rest of this document deals primarily with methods for numerically estimating the image function $I(x, y, \lambda)$ for a particular scene and camera position.

This work is a fairly holistic treatment of the image acquisition problem. It begins with fundamentals of optics and surface geometry which are well known in many fields, but are not as widely understood by Computer Graphics practitioners[2]. The machinery of light transport theory is then developed in the language of Illumination Engineering. These fundamentals are then used to justify, and in some cases modify, commonly used local illumination models. Image based methods, which approach image acquisition starting from the eye/film-plane, are then described. Particular attention is paid to image based methods that calculate global illumination effects. Some new ideas on sampling issues in image based methods are also described. Zonal methods, which calculate lighting effects in the scene in a view-independent manner are then described. The zonal method is also extended to allow for general reflectance types in a spatially and temporally efficient manner. The implementational issues involved in global illumination are then discussed. I have attempted to at least partially describe every stage of generating a computed image, and it is especially hoped that the important details of global illumination calculation are fully treated.

[2]This lack of foundational knowledge has been greatly alleviated since the publication of Roy Hall's book[43]

In Chapter 2, some relevant details of optics are presented, with a particular emphasis on surface appearance. The traditional arguments for assuming only the effects of Geometrical Optics are also given. Special attention is paid to the optical characteristics of pigmented surfaces such as paint.

Chapter 3 formulates the mathematics of light transport between surfaces. This yields the Rendering Equation of light transport. This particular formulation is more precise than previous formulations because care is taken here to insure that transparent surfaces such as glass are accounted for in a natural and unambiguous manner. The image generation problem is phrased in terms of an evluation of the Global Radiance Function, which specifies the directional spectral radiance at every point.

In Chapter 4, traditional computer graphics reflection models are examined using the principles of Chapters 2 and 3. A modification of the traditional model for polished surface reflection is also introduced. A new simplification in the model of diffuse transmission is also discussed.

Chapter 5 surveys *image space* methods of finding $I(x, y, \lambda)$. These methods find a solution by working from the image plane out into the scene. The stochastic sampling strategies of distributed ray tracing and path tracing are discussed in a rigorous manner. A method of reducing the number of shadow rays is presented; the collection of all light sources is viewed as one large light source with disconyinuities, and shadow rays are fired toward this 'one' source. This chapter also includes an analysis of static and adaptive sampling strategies, and introduces a new sampling strategy that has some desirable properties, and outperforms Poisson disk sampling for some images. The analysis introduces to computer graphics the concept of *discrepancy*, a measure from numerical integration theory that is useful in evaluating sampling strategies. Discrepancy is particularly useful because it associates a numerical quality estimate

to any particular sampling distribution. Readers unfamiliar with Monte Carlo integration may wish to look at Appendix B before reading Chapter 5.

Chapter 6 discusses zonal (radiosity) methods, and pays particular attention to formulating the progressive refinement methods as a physical simulation, rather than as a linear algebraic solution. This idea is used to develop a new zonal technique for non-diffuse surfaces. This chapter also includes a conjecture that we may be able to improve upon the $O(N^2)$ time complexity of traditional radiosity (N is the number of zones). To support this, a proof is given that, subject to certain constraints, a ray tracing solution to a N zone scene can be produced using only $O(N)$ rays. Assuming optimized ray intersection, this means a zonal solution can be produced in some envoronments in $O(N \log N)$ time.

Chapter 7 discusses zonal methods that account for participating media, such as smoke or fog, and a zonal method for arbitrary scattering properties is proposed. It is shown that the physical simulation strategy can be extended to participating media. This solution method has the advantage of being relatively inexpensive for sparse media. Light wavelength and time dependent problems, and methods of solving these problems are also discussed.

The implementational issues of global illumination simulations are discussed in Chapter 8. This discussion centers upon the C++ implementation used to produce the images in this work. It is shown that the local illumination model can be coded as a black box, and the global illumination code can be written without reference to the particular mix of local illumination models to be used.

Finally, Chapter 9 summarizes the work and lists several open questions.

CHAPTER 2

BASIC OPTICS

Almost every computer graphics program models optics in some way. Most make many simplifying assumptions about the physics of optics. In this chapter, the laws of basic optics are presented in a manner oriented toward graphics applications. In Section 2.1, the basic assumptions of physical and geometrical optics are outlined, and the visual impact of discarding physical optics is discussed. Section 2.2 covers the reflection of light from smooth, pure surfaces, primarily in the context of the Fresnel Equations. Section 2.3 discusses the reflection of light from roughened surfaces such as brushed steel. In Section 2.4 more complicated surfaces, such as paint, are described. Translucent materials, such as paper, are described in Section 2.5. Section 2.6 discusses clouds of small particles such as smoke and steam. Finally, the content of the chapter is summarized in Section 2.7.

Those readers with some knowledge of optics may safely skip this chapter (with the possible exception of Section 2.4).

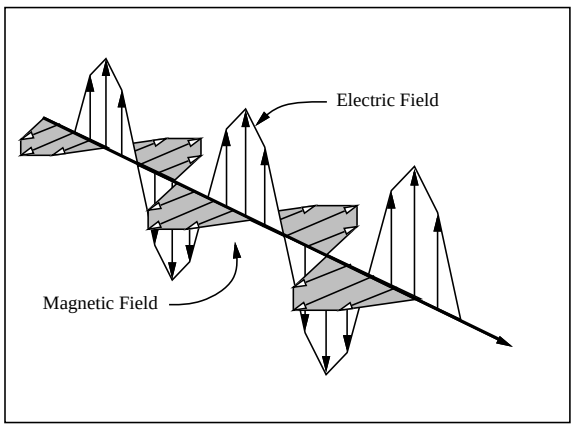

Figure 2.1: Propagation of an electromagnetic wave

2.1 Physical and Geometrical Optics

When trying to understand the behavior of light, physicists usually think of light either as propagating electromagnetic waves or as noninteracting energy carrying rays. The propagating waves have an oscillating electric field and an orthogonal oscillating magnetic field. The wave travels in a direction orthogonal to both fields as is shown in Figure 2.1. The average magnitude of the electric field associated with the wave determines the perceived energy of the wave. In the presence of two waves, the electric field values add (as vectors), so the two waves can interact in complicated ways that affect the perceived energy. If the particle model is used, the interactions are not so complicated because the photons are considered to be independent of each other. That is, the interactions are scalar rather than vector additions. The wave-based model of light behavior is called *physical optics* and the usually simpler ray based model is called *geometrical optics*.

In many situations physical optics effects are not visually important and we can think of the wave as just a ray with an energy (rather than vector field components), so that we can just sum the energy for two interacting rays.

In computer graphics, geometrical optics is almost always used instead of physical optics[1]. This section examines the impact of discarding physical optics in terms of what visible effects will be missing from computer generated pictures. Using geometrical optics allows light behavior to be modeled using many *independent* rays. The primary effects of physical optics that will be missing from pictures stem from interference, diffraction, and polarization.

Interference occurs when overlapping waves have their field values add such that cancellation or reenforcement takes place for a given wavelength. Visible examples of interference include:

- Colored rings in an oil film on a puddle.
- Colored bands on a soap bubble.
- Red and blue regions on a camera lens caused by antireflective coating.
- Bright colors of peacock feathers.
- Bright colors of mother of pearl.

Diffraction occurs when light waves bend traveling around a barrier. Diffraction effects are very important for long wavelength radiation, but are a fairly small effect for visible light. The primary type of visible diffraction is caused by *diffraction gratings*. Diffraction gratings are regular structures with spacings on the order of the wavelength of light. Some animals, such as certain beetles, have diffraction inducing structures that cause very bright colors.

Polarization effects are due to the orientation of the components of the electromagnetic field. For example, the glare of the sun off the pavement has a nonuniform distribution of orientation of field values. Polarized sunglasses preferentially filter these orientations. With the exception of such polarized filters, most examples (not involving glare effects) of visible polarization effects are fairly artificial. However, in some applications, particularly those that involve skylight, polarization can be important to appearance[110].

[1] A notable exception is the work by [75].

material	*refractive index*
Air	1.000293
Water	1.333
Diamond	2.419
Dense flint glass	1.75
Vitreous quartz (glass)	1.45
Salt	1.50

Table 2.1: Indices of refraction for various dielectrics (from Hecht and Zajac[49])

In computer graphics, geometrical optics is usually used because modeling light as waves rather than rays is very difficult, and because of the relative unimportance of the effects of interference, diffraction and polarization. When physical optics effects are modeled they are usually treated as special cases, as will be discussed in Chapter 7.

One physical optics parameter that is important to graphics is the wavelength of light. The wavelength distribution of a set of electromagnetic waves determines its color. This information can be included in geometrical optics by associating a wavelength (or wavelength distribution) with each ray.

2.2 Smooth Surfaces

Smooth surfaces of pure materials (such as smooth glass) reflect light in a well understood way described by the *Fresnel Equations* (Section 2.2.2). These equations predict the portion of light reflected and transmitted at a surface. The equations are usually only written for the interface between two dielectrics[2]. A dielectric is characterized by its index of refraction n. The index of refraction measures how much an electromagnetic wave slows down relative to its speed in vacuum. Typical values for n are listed in Table 2.1.

[2]Dielectrics are essentially non-metals (non-conductors). Semi-conductors are conductors with small extinction coefficient k. Most substances are conductors for at least some wavelengths.

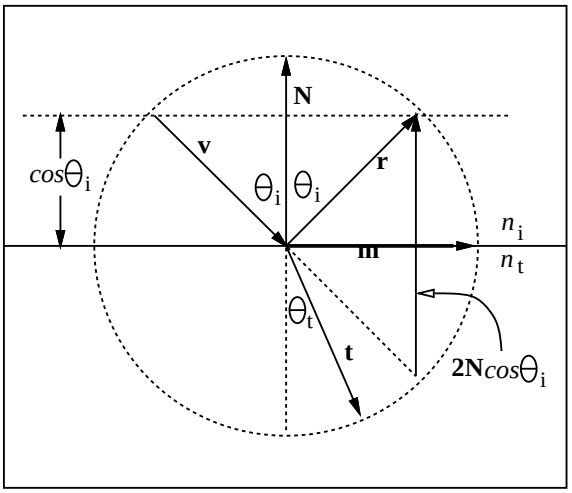

Figure 2.2: Geometry for light hitting a smooth surface.

2.2.1 Snell's Law

When light hits a smooth surface, a portion is reflected, and the rest is transmitted into the object. The direction of the reflected light is given by the Law of Reflection: the angle of the reflection direction $\mathbf{r}$, is equal to the angle of incidence θ_i, and will be in the same plane as the incident vector $\mathbf{v}$ and a surface normal vector $\mathbf{N}$. This arrangement is shown in Figure 2.2.

The direction of transmitted light, $\mathbf{t}$, is also in the same plane as $\mathbf{v}$ and $\mathbf{N}$, and obeys *Snell's Law*:

$$n_i \sin\theta_i = n_t \sin\theta_t \tag{2.1}$$

When Equation 2.1 cannot be satisfied, all of the light is reflected. Otherwise, a portion will be transmitted. The portions reflected and refracted can be determined using the Fresnel Equations described in the next section.

When implementing graphics simulations, it is convenient to determine $\mathbf{r}$ and $\mathbf{t}$ using vector arithmetic. The equation for $\mathbf{r}$ is fairly simple:

$$\mathbf{r} = \mathbf{v} + 2\mathbf{N}\cos\theta_i = \mathbf{v} - 2\mathbf{N}(\mathbf{v}\cdot\mathbf{N}) \tag{2.2}$$

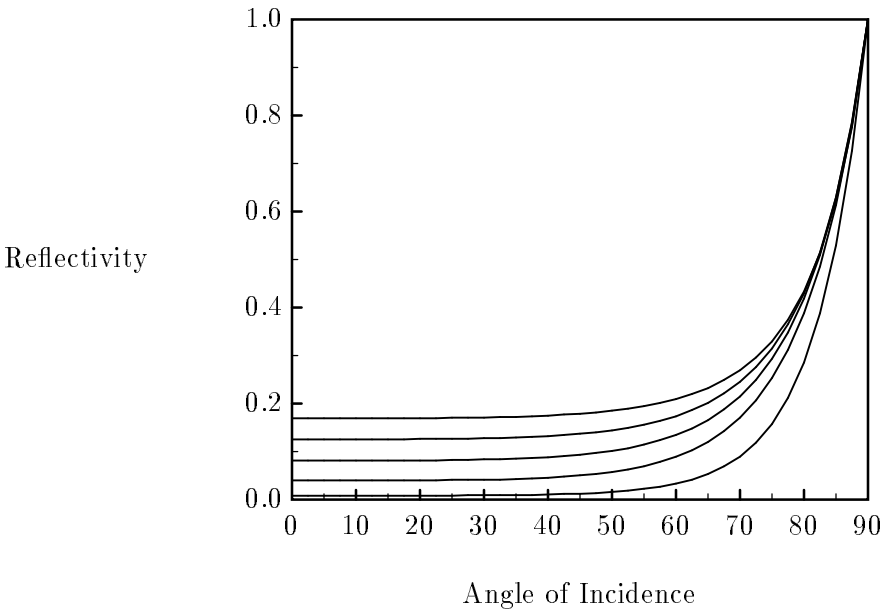

Figure 2.3: Reflectivity of light passing from air to dielectrics of various indices of refraction (1.2, 1.5, 1.8, 2.1, 2.4 bottom to top). Curves were generated directly from Fresnel Equations.

Finding $\mathbf{t}$ is a bit more tedious. From Figure 2.2, we can see that:

$$\mathbf{t} = -\mathbf{N}\cos\theta_t + \mathbf{m}\sin\theta_t \tag{2.3}$$

This eventually yields[52]:

$$\mathbf{t} = \left[-\frac{n_i}{n_t}(\mathbf{v}\cdot\mathbf{N}) - \sqrt{1-\left(\frac{n_i}{n_t}\right)^2(1-(\mathbf{v}\cdot\mathbf{N})^2)}\right]\mathbf{N} + \frac{n_i}{n_t}\mathbf{v}$$

2.2.2 Fresnel Equations

Conductors, or metals, are described by their refractive index and their extinction coefficient, k. The extinction coefficient of dielectrics is (by definition) zero. Given an interface between a dielectric with refractive index n_i, and a conductor with refractive index n and extinction coefficient k, the reflectivity can be found using the Fresnel Equations[96, 102]:

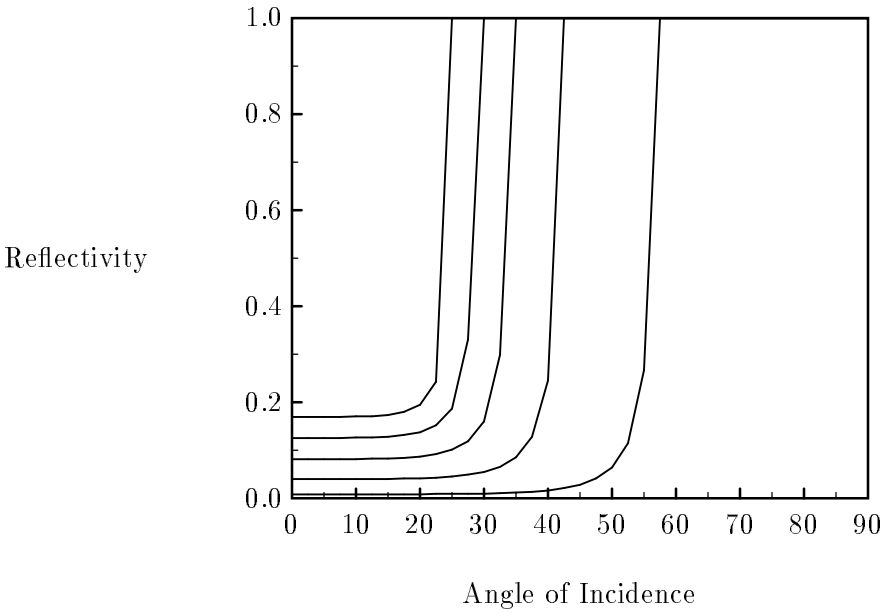

Figure 2.4: Reflectivity of light passing from dielectrics of various indices of refraction (1.2, 1.5, 1.8, 2.1, 2.4 bottom to top) to air. Curves were generated directly from Fresnel Equations.

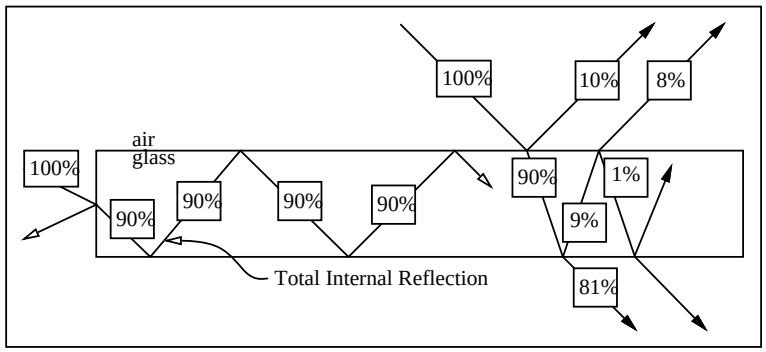

Figure 2.5: Total internal reflection (left) and branching reflection (right).

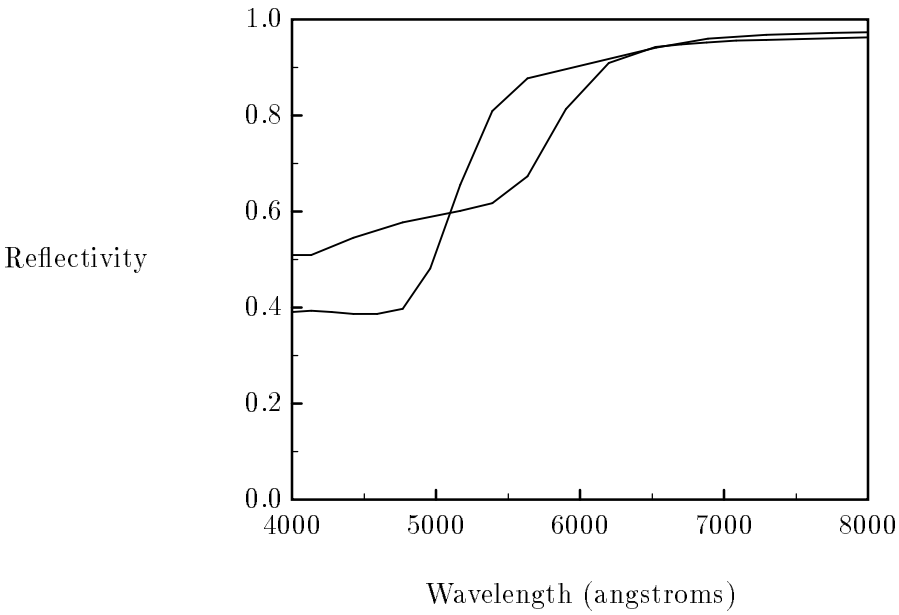

Figure 2.6: Normal incidence reflectivity for various wavelengths for gold (bottom curve at y-axis) and copper. Curves were generated directly from Fresnel Equations using data from Palik [82].

$$R_s = \frac{a^2 + b^2 - 2a\cos\theta + \cos^2\theta}{a^2 + b^2 + 2a\cos\theta + \cos^2\theta} \tag{2.4}$$

$$R_p = R_s \frac{a^2 + b^2 - 2a\sin\theta\tan\theta + \sin^2\theta\tan^2\theta}{a^2 + b^2 + 2a\sin\theta\tan\theta + \sin^2\theta\tan^2\theta} \tag{2.5}$$

Where R_s and R_p are the reflectivities for the two planes of polarization, θ is the angle of incidence, and a and b are given by

$$a^2 = \frac{1}{2n_i^2}\left\{\sqrt{(n^2 - k^2 - n_i^2\sin^2\theta)^2 + 4n^2k^2} + n^2 - k^2 - n_i^2\sin^2\theta\right\} \tag{2.6}$$

$$b^2 = \frac{1}{2n_i^2}\left\{\sqrt{(n^2 - k^2 - n_i^2\sin^2\theta)^2 + 4n^2k^2} - (n^2 - k^2 - n_i^2\sin^2\theta)\right\} \tag{2.7}$$

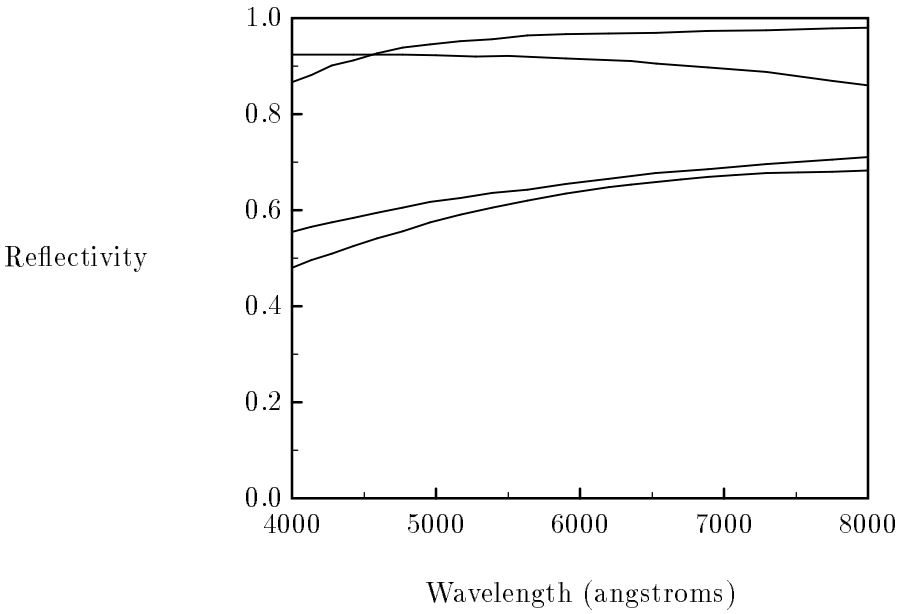

Figure 2.7: Normal incidence reflectivity for various wavelengths for (bottom to top on y-axis) nickel, platinum, silver, and aluminum. Curves were generated directly from Fresnel Equations using data from Palik [82].

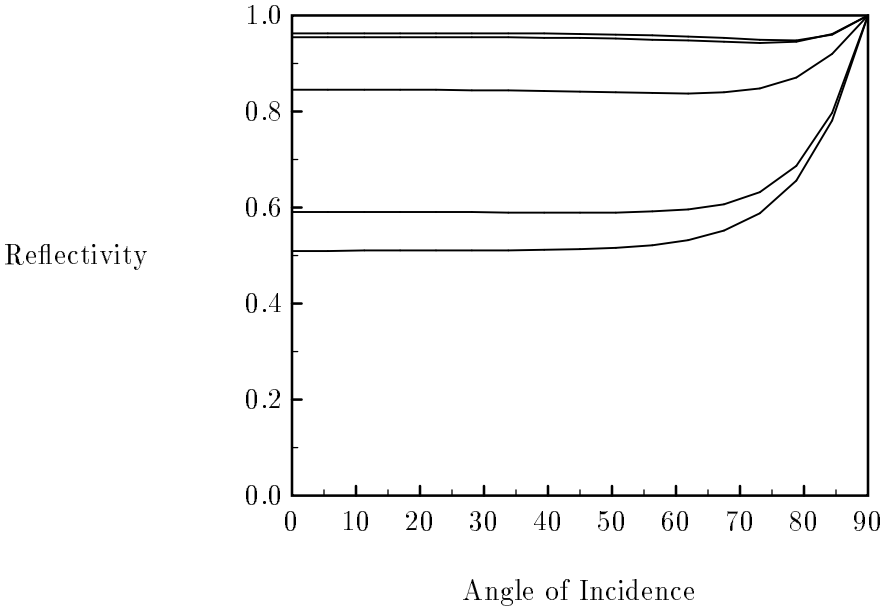

Figure 2.8: Reflectivity of copper for wavelengths 4000, 5000, 6000 7000, and 8000 angstroms (bottom to top). Curves were generated directly from Fresnel Equations using data from Palik [82].

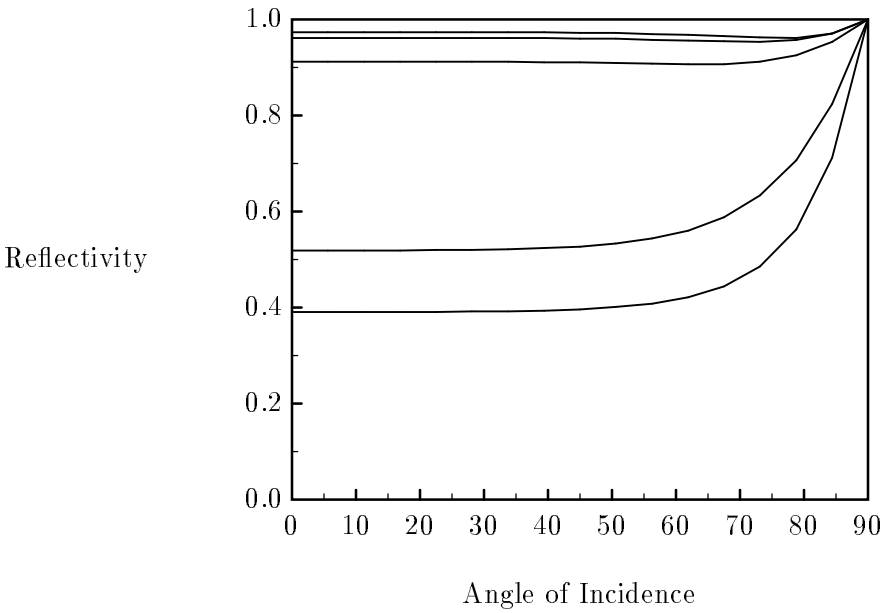

Figure 2.9: Reflectivity of gold for wavelengths 4000, 5000, 6000 7000, and 8000 angstroms (bottom to top). Curves were generated directly from Fresnel Equations using data from Palik [82].

The reflectance for unpolarized light is simply the average of R_s and R_p. The equations for a dielectric-dielectric interface can be found by setting k to zero and using n as the index of refraction for the second dielectric.

Values of n and k can vary with wavelength, so the color properties of materials are treated automatically. Values of n and k for many materials at optical and other wavelengths can be found in Palik [82].

There is an important property of the equations that should be mentioned. First, when θ is 90°, both R_s and R_p go to one for all physical values of the optical constants[3]. This is why we see glare off surfaces when we see them at an acute angle. Another property of the equations is

[3]To see that R_p goes to one, multiply R_p by $\cos^2\theta/cos^2\theta$, or employ l'Hopital's Rule.

that they can be employed without regard to the direction of propagation. This property will become important when ray tracing methods are discussed in Chapter 5.

The reflectivity at an air-dielectric interface is shown in Figures 2.3 and 2.4. Notice in Figure 2.4 that light passing from a dielectric with a higher index of refraction to air will at some angle of incidence reflect completely. This phenomenon is called *total internal reflection*, an example of which is shown on the left of Figure 2.5. Because of total internal reflection, light can travel inside a transparent object for many reflections with almost no attenuation, as happens in fiber optic cables. This property will have an important bearing on the ray tree pruning rules described in Chapter 5.

The refractive index of a dielectric will typically vary somewhat with wavelength, so the reflectivity and angle of transmission can vary for different colors of light. This is why a prism is able to separate a beam of white light into a spectrum. This divergence of transmitted light is called *dispersion*, and will be discussed at greater length in Chapter 7. For metals, both n and k can vary with wavelength. This is why metals can take on 'non-silver' colors. The reflectance curves for two such metals, gold and copper, are shown in Figure 2.6 for normal incidence ($\theta = 0$). The reflectance curves for several 'silver' metals are shown in Figure 2.7. As evident in the figure, these metals are also colored, but not to the extremes of gold or copper. The angular behaviors of the reflectance of gold and copper for several wavelengths are shown in Figures 2.8 and 2.9. As expected, the reflectivity at $\theta = 90°$ goes to one for all wavelengths.

2.2.3 Absorption of Transmitted Light

Once light is transmitted into a medium, it may be absorbed. In a dielectric, this will happen if there are absorptive molecules, such as dye molecules, inside the medium. Dielectrics with

an isotropic distribution of absorptive molecules, such as green glass, obey *Beer's Law*:

$$I(\lambda) = I_0(\lambda)e^{-C(\lambda)t}$$

where $I_0(\lambda)$ is the intensity of the light at wavelength λ just after passing into the dielectric, $I(\lambda)$ is the intensity after passing through a distance t of the dielectric, and $C(\lambda)$ is a positive constant that describes the behavior of the dye.

Metals also absorb transmitted light, but they do it without dyes, and do it so quickly that they are opaque for almost any thickness. If the thickness, t, is very small, such as the thin coating of gold film on some sunglasses, then a metal can be partially transparent. The degree of transparency is determined by *Lambert's Law*[4]:

$$I(\lambda) = I_0(\lambda)e^{-4\pi kt/\lambda}$$

where k is the extinction coefficient of the metal at wavelength λ. Lambert's Law can, of course, be thought of as a special case of Beer's Law, where C is based solely on the character of the metal. If there are two adjacent thin films of metal, Lambert's Law can still be used, but a more general form of the Fresnel Equations would be needed to determine boundary reflectivity.

2.3 Rough Surfaces

Most surfaces in real environments are not smooth and of pure composition, so we cannot directly apply the Fresnel Equations to determine a surface's reflective behavior. A great body of both the Heat Transferliterature[106, 102, 80], and the Computer Graphics literature[27, 60, 16, 111] has been devoted to the case when the surface is not smooth, but is made up of a pure material. A close-up view of such a surface is shown in Figure 2.10. Examples of such surfaces

[4]Not to be confused with Lambert's Law of diffuse reflection.

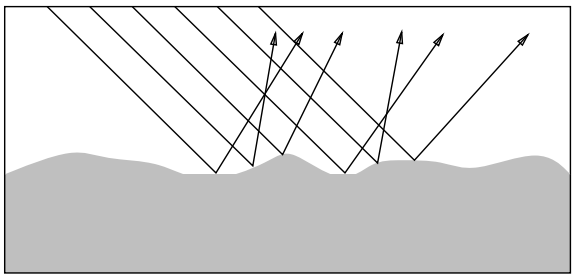

Figure 2.10: Detailed view of a rough surface.

include etched glass and brushed steel. Different ways of treating these surfaces for Computer Graphics purposes are detailed in Chapter 4.

2.4 Composite Surfaces

Most surfaces in nature, and even in man-made environments, are neither smooth nor of pure composition. These *composite* surfaces are made up of at at least two materials in some interesting arrangement. The models of composite surfaces discussed in this section are based largely on the discussion in the books by Rossotti[89], Falk[33], and Williams[122]. More realistic models of composite surfaces allow for varied particle shape and nonuniform particle distribution. Discussion of the subtleties of these general surfaces can be found in Parfitt and Sing[83], and Feller[34].

A classic example of a composite surface is a gloss paint which is made up of some kind of dielectric medium with suspended pigment particles, as shown in Figure 2.11. The pigment particles for a white paint are typically clear dielectrics or 'silver' conductors, with a different index of refraction from the medium. The greater the index of refraction, the greater is the quantity of light scattered at the medium-pigment interface, and the more opaque the paint. Some paints have a medium whose index of refraction changes as it dries, which explains the

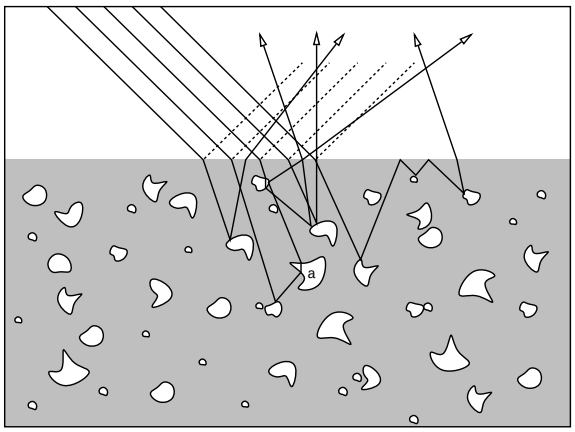

Figure 2.11: Detailed view of a glossy paint.

loss of opacity for some drying paints. In a sense, snow is also a composite surface made up of ice and air. The geometry of the snow flakes on the ground can produce complex reflective behavior[73]. Another good example is an ice cubes with internal air bubbles. If there are many bubbles, the ice becomes opaque; the ice is the medium, and the air bubbles are the pigment particles.

If the pigment particles have a color (caused by dye or positive k), then there will be selective absorption for light traveling through the pigment (such as the light passing through particle 'a' in Figure 2.11), and the paint will be colored. Because the surface of the paint reflects light according to the Fresnel Equations, it will behave as a partial mirror. This mirrorlike behavior accounts for the highlights and reflections seen on gloss painted surfaces.

Many walls are painted with a paint roller or spraygun and have a rough surface. Such surfaces still exhibit mirrorlike behavior, as shown in Figure 2.12, but the reflections may lack coherence, so the highlights will be spread out. Some paints are so rough that they do not look reflective; they are *matte*. Another way paint can look matte is to have the pigment particles pierce the surface of the medium, either by wear or design, as shown in Figure 2.13.

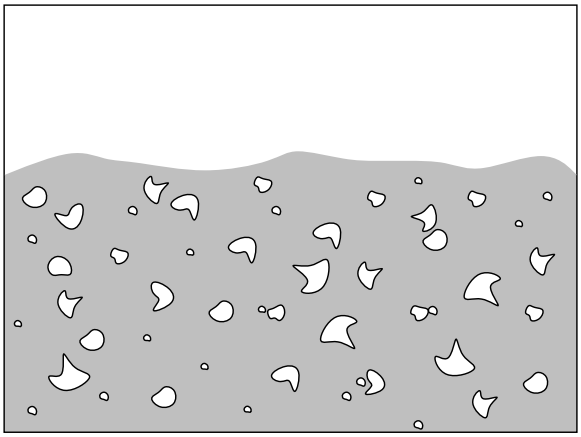

Figure 2.12: Detailed view of a glossy paint with uneven surface.

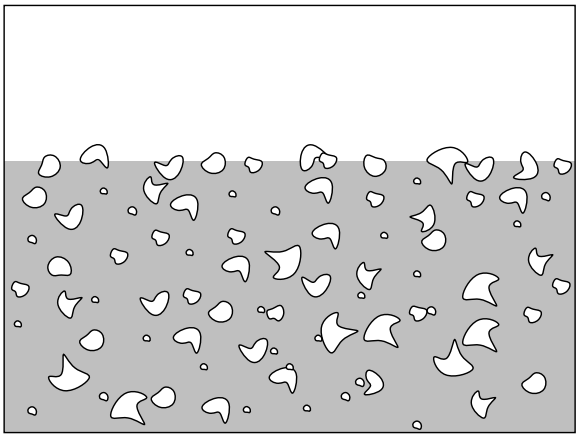

Figure 2.13: Detailed view of a matte paint

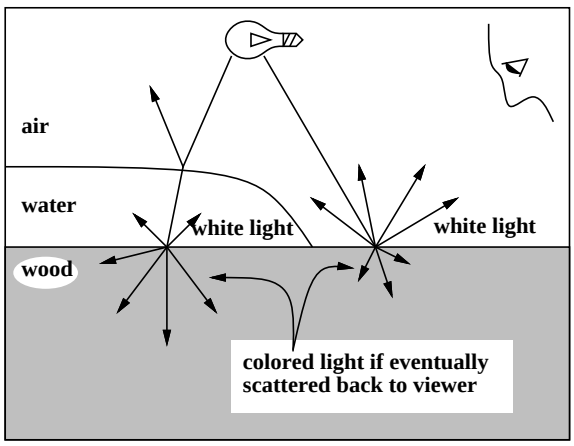

Figure 2.14: Color of wet wood is deeper than color of dry wood

One interesting thing about a polished surface is that its color is usually darker and more saturated than an unpolished surface of the same material. For example, most dark hardwoods get a much deeper color when wet or varnished. This is because the wood particles have an index of refraction that is closer to that of the water than that of air[89]. This means more light will be transmitted from water to the wood than from air to wood, and less light will be reflected at the wood surface. The transmitted light that eventually scatters back out of the wood will have been filtered to be 'wood colored', as opposed to light reflected uniformly across wavelengths at the surface. The light reflected at the water-air interface is reflected specularly (the surface is smooth), so this light is not seen by the viewer (unless the viewer is in the path of reflection, in which case a bright highlight will be seen). This idea is illustrated in Figure 2.14.

2.5 Translucent Surfaces

Some rough surfaces, such as paper, allow some light to pass through. These *translucent* objects are made up of many small particles or fibers that scatter light. They are translucent if their thickness is small enough that a significant amount of light can make it from one side of the

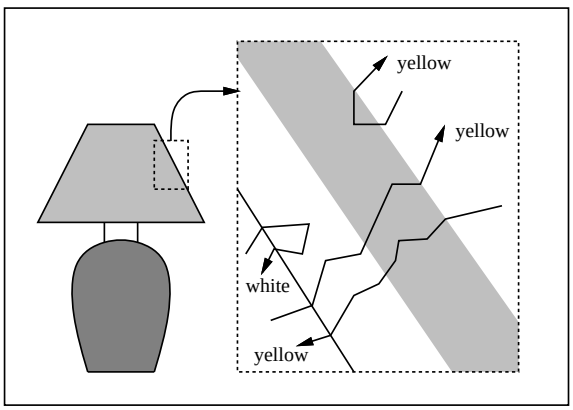

Figure 2.15: Detailed view of a translucent lampshade.

object to the other. As an example, consider a pile of ground glass on a sheet of red paper. Near the center of the pile the glass will appear white because almost all of the light will be backscattered out of the pile before reaching the paper. Near the edges of the pile, where there is less glass to scatter the light, some light will reflect off the red paper, and the red color will show through. This red part of the pile is translucent. The rest of the pile is opaque. Of course, all of the pile will allow *some* portion of incident light to reach the red paper, so opacity is a matter of degree.

A detailed view of a translucent paper lampshade is shown in Figure 2.15. The paper has different reflective properties on the two sides, but should have roughly the same transmittance properties for each side. This is because of the Helmhotz reciprocity rule: The transmittance is a function of the available paths through the shade, and these paths can be traveled in either direction.

2.6 Participating Media

Whenever the environment to be modeled is not associated with a vacuum, the region between surfaces will be occupied by some kind of atmosphere such as air or water. This atmosphere is

made up of many molecular particles which can scatter light, and may include larger particles such as dirt, smoke, or steam that also scatter light. A discussion of such effects from a computer graphics perspective, and an excellent treatment of scattering in general, can be found in the paper by Klassen[67]. A more formal treatment can be found in the classic book by van de Hulst[116].

The scattering by large particles can be looked at in terms of geometrical optics. If the particles are small relative to the wavelength of light, then physical optics predicts that light will scatter as if the particle had a larger size. This type of scattering is called *Raleigh Scattering* and is highly biased toward short (blue) wavelengths. Raleigh Scattering is responsible for the blue color of the sky and the color of blue eyes (the eyes have many small particles in suspension). In between classical and Raleigh scattering both physical and geometrical optics are present. This combined type of scattering is called *Mie Scattering*, and the associated theory is complex. All of the scattering can be predicted by associating with each type of particle a scattering cross section and a scattering distribution function.

2.7 Summary

Light can be considered to be rays (geometrical optics), or waves (physical optics). Geometrical optics is usually used in computer graphics because of the independence property of rays. Ignoring physical optics will preclude modeling some observable effects such as interference and diffraction.

Smooth surfaces reflect and transmit light in a well understood way governed by the Fresnel Equations. The equations can be applied when the refractive index and extinction coefficient of a material are known.

Surfaces that are not macroscopically smooth can be better understood by looking at their microscopic structure. Rough surfaces appear matte because of their distributed surface normals. Polished surfaces exhibit mirrorlike behavior at their surface, and materials in or under the polish can scatter the transmitted light in a non-specular way. Thin objects made up of particles or fibers can allow some light to pass through the object, so they exhibit translucent behavior.

Small particles suspended in space can scatter light between surfaces. If the wavelength of light is small relative to the size of the particles, then the scattering can be predicted using geometrical optics. If the wavelength is large relative to particle size, then Raleigh Scattering theory can be used. For intermediate wavelengths, Mie Scattering theory must be used.