Predicting Reflectance Functions from Complex Surfaces

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Stephen H. Westin

August 1992

ABSTRACT

This thesis describes a physically-based Monte Carlo technique for approximating bidirectional reflectance distribution functions (BRDFs) for a large class of geometries by directly simulating geometric optical scattering from surfaces. The method is more general than previous analytical models: it removes most restrictions on surface microgeometry. Three main points are described: a new representation of the BRDF, a Monte Carlo technique to estimate the coefficients of the representation, and the means of creating a milliscale BRDF from microscale scattering events. The combination of these techniques allows the prediction of scattering from essentially arbitrary roughness geometries. The BRDF is concisely represented by a matrix of spherical harmonic coefficients; the matrix is directly estimated from a geometric optics simulation, enforcing exact reciprocity. Microscale scattering events are represented by direct simulation (e.g., specular reflection and transmission by individual textile fibers) or by a microscale-averaged model (e.g., a wave-optics-based statistical BRDF) depending on the ratio of the wavelength of light to the size of microscale surface features. These microscale events are integrated over some area of larger-scale geometry to obtain a milliscale BRDF for the overall surface. The process may be repeated for larger scales of spatial integration. The method applies to roughness and feature scales that are much larger than the wavelength of light, yet much smaller than the spatial density at which the BRDF is sampled across the surface, say, for image generation. Example surfaces include brushed metal and textiles. The method is validated by comparing with exact solutions and with an existing scattering model, and sample images are generated with a physically-based global illumination algorithm.

Biographical Sketch

Stephen H. Westin was born September 21, 1955 in Flint, Michigan. He was granted the degree Bachelor of Science in Computer Engineering from the University of Michigan, Ann Arbor, in 1980. He was employed as a system programmer and system manager at the Environmental Research Institute of Michigan, Ann Arbor, from 1976 to 1982. He then provided software support for structural analysis at Ford Motor Company, Dearborn, Michigan until 1984, when he joined Ford Design Staff in Dearborn to help plan, implement, and support computer graphics tools for aesthetic design. He was assigned to the Ford Design Centre in Köln, Germany from 1987 to 1990. Upon his return from Germany, he entered the Program of Computer Graphics at Cornell to pursue a Master's degree.

Acknowledgements

First, many thanks to the best office mate and co-author in the world, Jim Arvo. None of this would have happened without his insight, ideas and encouragement. Thanks also go to Ken Torrance, who always knew the difference between energy and intensity, and was patient and lucid in explaining it. Thanks to Don Greenberg, the driving force behind the Program of Computer Graphics, for taking a chance on me and helping to guide this research. Ford Motor Company Design Staff funded a fellowship which supported me during my time at Cornell. Special thanks go to my wife, Barbara, and to Peter and Grace for their involuntary donation of so much of my time while working on this thesis.

Thanks also to:

- François, who let me get involved
- Ben, the keeper of the Rendering Testbed
- Hurf, who kept the machines running
- Julie and Kurk, who helped when it mattered
- David Field, who really hasn't been forgotten
- Suzanne, who helped with the diagrams
- Brian, who always asked intelligent questions

- Harold, the LaTeXMan
- Sabine Coquillart, who provided the geometric data for the cushion in Figures 6.17 and 6.20
- Barry, Giuseppe, and Dave
- the Apollo DN10000 design team, wherever you are

Table of Contents

List of Figures

Chapter 1

Introduction

How can we predict the direction and color of light scattering from a surface? Since the earliest days of computer graphics, experimenters have recognized that an image is no more realistic than the model of local light scattering used to make it [Gouraud 71,Phong 75].

Models of local scattering in computer graphics started with *ideal diffuse*, i.e. Lambertian, models [Appel 68,Wylie 67], but Phong soon added an ad-hoc model of *directional-diffuse* scattering [Phong 75] and Whitted added *ideal specular* (mirror-like) reflection [Whitted 80]. While simple to compute, these models exclude many important physical effects and lack the energy consistency needed for global illumination calculations.

Blinn [Blinn 77] and others [Cook 82,He 91] built on earlier work in radiative transfer [Torrance 67] to introduce progressively more sophisticated, physically-based, analytical scattering models. These predict the *bidirectional reflection func-*

tion (BRDF) with good fidelity to physical measurements, but apply only to surfaces of specific types of random isotropic roughness.

Several authors have proposed scattering models for anisotropic surfaces. Poulin and Fournier [Poulin 90] developed a closed-form solution for a surface of parallel cylinders or cylindrical grooves, and Kajiya [Kajiya 85] explored a numerical technique using Fourier coefficients and bump maps. Neither of these models guarantees physical consistency (e.g. energy conservation). Each method is restricted to certain characterizations of surface roughness; the accuracy of each model depends on how well this description matches the physical surface to be modeled.

This thesis presents a new method of creating local scattering models. The method has three main components: a concise, general representation of the BRDF in terms of spherical harmonics over the incident and reflected hemispheres, a Monte Carlo technique to estimate the coefficients of the general representation, and a means of using scattering at one scale to create a BRDF for a larger scale. The method uses geometric optical scattering to model rough surface effects (surface feature size $\gg \lambda$ where λ is the wavelength of light), using an analytical BRDF model to include smooth surface effects (surface feature size $\approx \lambda$), when wave phenomena must be included. This allows the prediction of scattering from essentially arbitrary geometries, accurately modeling both isotropic and anisotropic materials such as brushed metals, velvet, and woven textiles.

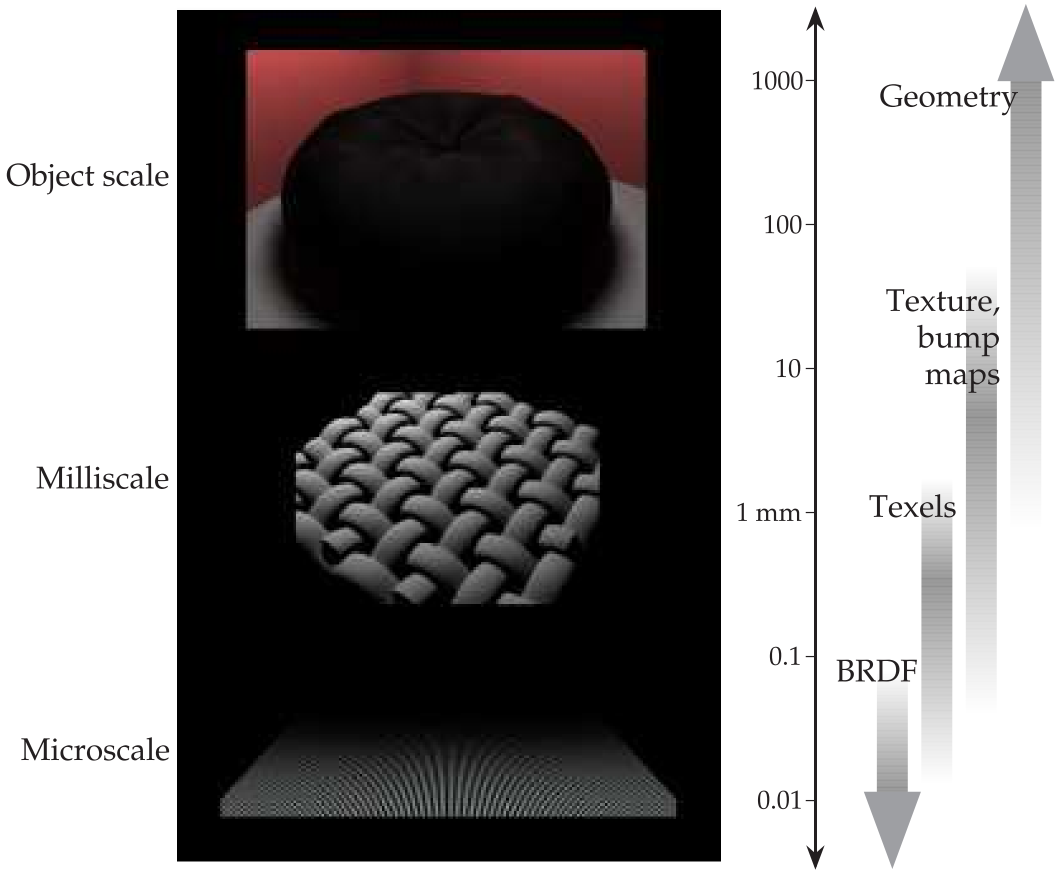

Figure 1.1: Scales of Applicability

The approach outlined in this thesis is applicable wherever the BRDF is an adequate model of surface geometry. A local scattering model represents the smallest scales of geometry in an environment (size $\ll$ 1 mm); we call this level *microgeometry* (Figure 1.1). At the scale of *milligeometry*, i.e. features on the scale of about 1*mm*, we often use texture maps, bump maps, or texels; at the largest scale, *object scale*, it is necessary to model geometry explicitly, for example with polygons or parametric patches. The scales of applicability of each method depend on the result desired: the upper limit of applicable scale is determined by the frequency of sampling across a surface, and the lower limit is determined by the wavelength λ of radiation being modeled.

Scale ultimately depends on the Nyquist limit, the highest frequency which can be reconstructed from discrete samples [Oppenheim 89]. In rendering, no feature smaller than about twice pixel size need be represented explicitly. This size obviously varies with point of view. When rendering, say, an interior scene, objects as small as a pencil must be modeled at object scale; when simulating the view from miles away, however, objects as large as trees and buildings can be modeled as local scattering, so we can think of them as microgeometry.

The advent of global illumination methods (e.g. radiosity [Goral 84] and enhanced ray tracing [Ward 88]) has created another concept of scale by modelling the transfer of light between objects in the environment, not just from each object to the eye. To do this, the illumination is sampled in object space, with far fewer samples than are used for rendering. This lower sampling frequency reduces the Nyquist limit, so many objects which are visible at the image sample frequency ($N \approx 10^5$) fall below the limit of representation at the much lower illumination sample frequency ($N \approx 10^6$). These small objects can be treated as microgeometry in solving the global energy balance, though they are explicitly represented for rendering. Global illumination methods thus can rely on the BRDF over a much wider range of scales, simplifying the explicit geometry used. These methods demand an energy-consistent BRDF, however, as any inaccuracies can be multiplied through many bounces.

The thesis starts with a description of the physics of geometric optical scattering, then presents the major components of the technique: the BRDF representation, the Monte Carlo estimator, and the means of estimating a milliscale BRDF from the microscale description of surface roughness. Finally, several examples of applying the method are shown.

Chapter 2

Basic Geometric Optical Scattering

2.1 Geometry of Scattering

When light strikes an object, generally some portion will be reflected, some portion will be transmitted, and some portion will be absorbed. In computer graphics, absorption is of no interest; only reflected or transmitted light can reach the eye of the observer.

Optical scattering from an opaque surface is defined by the geometry in Figure 2.1. Light strikes a surface from an incident direction (θ_i, ϕ_i) within some solid angle $d\omega_i$ and is scattered in a reflection direction (θ_r, ϕ_r) over some solid angle $d\omega_r$. The coordinate system is aligned so that the z axis is perpendicular to the surface.

This thesis deals primarily with *geometric scattering*, where light is assumed to consist of particles that follow straight paths through a homogeneous medium

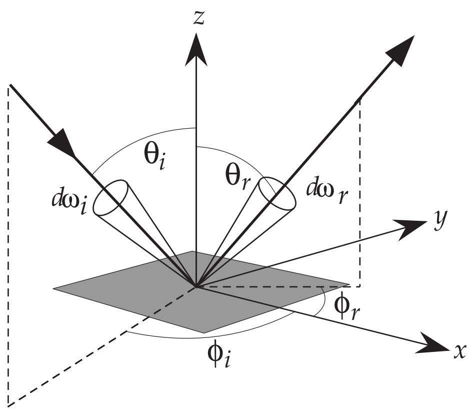

Figure 2.1: Geometry of Scattering

and do not interfere with each other. When all scattering features are much larger than the wavelength of light used, this assumption is an excellent approximation to the actual propagation of light. When the size of geometric features of the scattering surfaces approaches the wavelength of light, the light must be treated as waves that can diffract and interfere with each other.

2.2 Scattering Modes

Reflective scattering can be divided into three components, called *ideal specular*, *ideal diffuse*, and *directional diffuse* [He 91]. These modes are illustrated in Figure 2.2. Ideal specular reflection is a Dirac delta function: zero everywhere except in the specular direction $\theta_r = \theta_i, \phi_r = \phi_i + \pi$. This is a mirror-like reflection. Ideal diffuse scattering, also known as Lambertian scattering, is independent of reflection direction. Directional diffuse scattering is everything other than ideal specular or ideal diffuse scattering; it is the most complex scattering mode. A

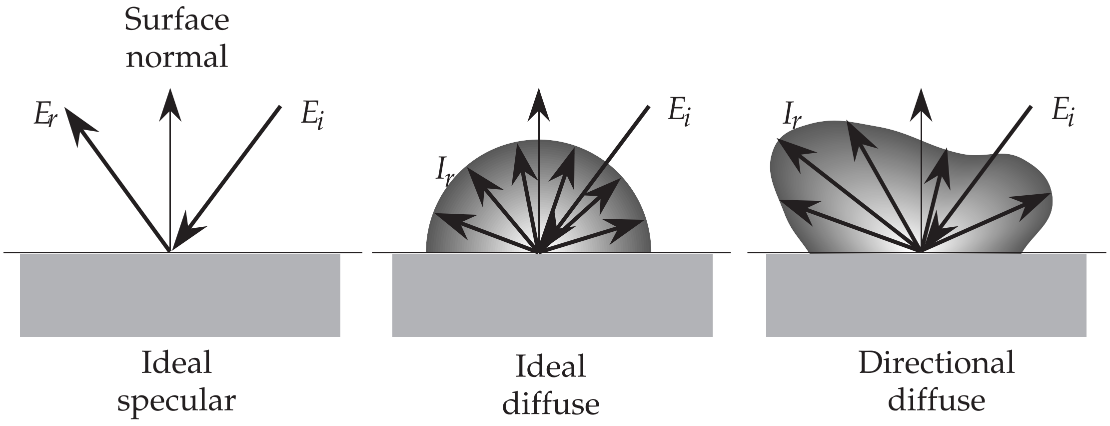

Figure 2.2: Local Scattering Modes

given surface may approach one of the pure modes shown in Figure 2.2, or its reflection behavior may be a combination of the three modes.

2.2.1 Ideal Specular Reflection

A perfectly smooth surface, or one which is only slightly rough, will reflect some portion of the incident beam in the specular direction $\theta_s = \theta_i, \phi_s = \phi_i + \pi$. This specular reflection preserves the solid angle of the incident beam (i.e. $d\omega_r = d\omega_i$), rather than spreading the reflected beam over the hemisphere, as with ideal diffuse or directional diffuse scattering. Reflection of unpolarized light from a perfectly smooth surface of a dielectric (i.e. non-conductive) material is governed by the Fresnel equation [Hecht 87]:

$$\begin{aligned} E_r &= F_\perp^2 + F_\parallel^2 \\ &= E_i \left(\frac{n_i \cos\theta_i - n_t \cos\theta_t}{n_i \cos\theta_i + n_t \cos\theta_t} \right)^2 + \left(\frac{n_t \cos\theta_i - n_i \cos\theta_t}{n_i \cos\theta_t + n_t \cos\theta_i} \right)^2 \end{aligned} \tag{2.1}$$

where E_r and E_i are the reflected and incident energy flux densities (defined in Section 2.3.1), $F_\perp$ and $F_{||}$ are the Fresnel reflectivities for light polarized perpendicular ($\perp$) and parallel ($||$) to the plane of incidence (shown in Figure 2.1), and n_i and n_t are refractive indices. A similar, but much more complex, expression governs reflection from a conductive surface. Any roughness of the surface at a scale comparable to or larger than the wavelength of light will alter this dependence on incident angle. He [He 91] has calculated the effects of wave optics and the dependence of effective roughness on incident angle to derive this relation for surfaces whose roughness has a Gaussian autocorrelation function. Specular scattering results from a single reflection from the surface [He 91]; it takes place only at the boundary between two materials of differing dielectric constants.

2.2.2 Ideal Diffuse Reflection

Ideal diffuse reflection is independent of reflected direction and depends only on incident direction. It is also known as Lambertian reflection.

$$I_r = k \cos \theta_i I_i(\theta_i, \phi_i) \tag{2.2}$$

where I_r and I_i are the reflected and incident radiances (defined in Section 2.3.2) and k is a constant. This scattering can result from multiple scattering events, either from a rough surface or from subsurface scattering. Either the incident beam is scattered many times from a rough surface, or it enters a dielectric, is scattered many times by particles beneath the surface, and is scattered back out.

2.2.3 Directional Diffuse Reflection

The most complex transfer is directional diffuse scattering, which includes all scattering which is neither ideal diffuse nor ideal specular. The scattering pattern can take on almost arbitrary, irregular shapes [Torrance 66]. This scattering results mainly from single scattering from a rough surface. Multiple scattering from the surface can also contribute to this component.

2.3 Definitions

2.3.1 Energy Flux Density

Energy flux density dE is the power received, emitted, transmitted, or reflected per unit area of a surface, $\frac{W}{m^2}$. It applies either to a finite wavelength band of light or for a given wavelength. In the latter case, the units become $\frac{W}{m^3}$.

2.3.2 Radiance

Radiance I is the energy flux per unit projected area per unit solid angle (e.g. $\frac{W}{m^2 sr}$)

$$I \equiv \frac{d^2\Phi}{dA_{proj}d\omega} = \frac{d^2\Phi}{dA\cos\theta d\omega} = \frac{dE}{\cos\theta d\omega} \tag{2.3}$$

where Φ is the radiant power (in watts), A_{proj} is the projected area in the direction of propagation and A is the surface area over which this energy is distributed. The angle θ may either θ_i or θ_r in Figure 2.1, depending on whether incident or reflected radiance is being described. The factor $\cos\theta$ accounts for the dependence of projected surface area on θ. Radiance is preserved along any path of propagation through an ideally transparent medium (e.g. vacuum), since the power Φ in the numerator declines as $\frac{1}{r^2}$, and so does the solid angle $d\omega$ in the denominator. Energy flux density can be calculated from radiance by

$$\begin{aligned} dE(\theta,\phi) &= \frac{d^2\Phi(\theta,\phi)}{dA} \\ &= I(\theta,\phi)\cos\theta d\omega. \end{aligned} \tag{2.4}$$

2.3.3 BRDF

Diffuse reflective scattering is completely characterized by the *bidirectional reflectance distribution function* (BRDF) $\rho_{bd}(\theta_i,\phi_i,\theta_r,\phi_r)$ [Nicodemus 77]. This is the radiance dI_r leaving the surface in direction (θ_r,ϕ_r) as a function of the energy flux density dE_i incident on the surface.

$$\rho_{bd}(\theta_i,\phi_i,\theta_r,\phi_r) = \frac{dI_r(\theta_r,\phi_r)}{dE_i(\theta_i,\phi_i)} \tag{2.5}$$

The function ρ_{bd} is called *bidirectional* because it is a function of both the incident and reflected directions; it obeys the law of *reciprocity* [Siegel 81, p. 65], i.e.

$$\rho_{bd}(\theta_i,\phi_i,\theta_r,\phi_r) = \rho_{bd}(\theta_r,\phi_r,\theta_i,\phi_i) \tag{2.6}$$

for all angles $(\theta_r,\phi_r,\theta_i,\phi_i)$. It can vary in magnitude from 0 to ∞.

The radiance in a scattering direction (θ_r,ϕ_r) is the inner product of the BRDF ρ_{bd} with the illumination $dE_i(\theta_i,\phi_i)$ over the entire incident hemisphere.

$$\begin{aligned} I_r(\theta_r,\phi_r) &= \int_{\Theta} \rho_{bd}(\theta_i,\phi_i,\theta_r,\phi_r) dE_i(\theta_i,\phi_i) \\ &= \langle \rho_{bd}(\theta_i,\phi_i,\theta_r,\phi_r) \,|\, E_i(\theta_i,\phi_i) \rangle \end{aligned} \tag{2.7}$$

where the symbol $\frown$ denotes integration over the hemisphere.

2.3.4 Directional Hemispherical Reflectance ρ_{dh}

The directional hemispherical reflectance $\rho_{dh}(\theta_i, \phi_i)$ is the integral of $\rho_{bd} \cos \theta_r$ over the entire reflection hemisphere [Siegel 81, pp. 66-67].

$$\rho_{dh}(\theta_i, \phi_i) \equiv \int_{\frown} \rho_{bd} \cos \theta_r d\omega_r \tag{2.8}$$

It is the total reflected energy flux density per unit incident energy flux density, and can never exceed unity.

2.3.5 Hemispherical Directional Reflectance ρ_{hd}

The reciprocal quantity associated with ρ_{dh} is $\rho_{hd}(\theta_r, \phi_r)$, the reflectance in a given direction (θ_r, ϕ_r) resulting from illumination over the hemisphere. It is the ratio of reflected radiance in the direction (θ, ϕ) to the average incident radiance. If the incident radiance is uniform over the incident hemisphere, the hemispherical reflectances satisfy the reciprocity relationship

$$\rho_{dh}(\theta, \phi) = \rho_{dh}(\theta, \phi). \tag{2.9}$$

2.3.6 Specular Reflectance ρ_s

Ideal specular reflection is characterized by the specular reflectivity

$$\rho_s = \frac{dE_r}{dE_i}. \tag{2.10}$$

Since the solid angle of the beam is preserved, the relation also holds for radiance:

$$\rho_s = \frac{I_r}{I_i}. \tag{2.11}$$

The specular reflectance is related to the bidirectional reflectance ρ_{bd} by

$$\rho_s(\theta_i, \phi_i) = \rho_{bd}(\theta_i, \phi_i, \theta_r = \theta_i, \phi_r = \phi_i + \pi) \cos\theta_i d\omega_i \tag{2.12}$$

or

$$\rho_{bd}(\theta_i, \phi_i, \theta_r = \theta_i, \phi_r = \phi_i + \pi) = \frac{\rho_s(\theta_i, \phi_i)}{cos\theta_i d\omega_i}. \tag{2.13}$$

For specular surfaces, those with $0 < \rho_s \leq 1$, the bidirectional reflectance increases without bound as $d\omega \to 0$.

Chapter 3

Wheels Within Wheels: Representing the BRDF with Spherical Harmonics

A general scattering function for unpolarized light is a function of four variables, $\rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r) : S^2 \times S^2 \mapsto \Re$, where S^2 is the unit sphere, θ_i, ϕ_i are the elevation and azimuth angles of incidence, and θ_r, ϕ_r are the corresponding angles of reflection. For a BRDF, ρ_{bd} is zero whenever θ_i or $\theta_r > \frac{\pi}{2}$. Since the BRDF can take on highly arbitrary shapes [He 91,Torrance 66], a very general method is needed to represent it. Fortunately, a BRDF is generally smooth, making it a good candidate for representation by smooth orthogonal functions. Previous authors have used spherical harmonics to represent scattering functions [Cabral 87,Kajiya 84, Sillion 91], since they form a complete basis set of smooth functions over the sphere. Kajiya [Kajiya 84] used spherical harmonics to derive an analytical scattering function; Cabral et al. [Cabral 87] and Sillion et al. [Sillion 91] used them as a numerical approximation to the BRDF. The representation used in this thesis is an extension of Sillion's technique; it provides an accurate, concise embodiment of the general BRDF.

3.1 Overview of Spherical Harmonics

Any square-integrable function over the sphere can be exactly represented by an infinite sum of spherical harmonic basis functions, $Y_{lm}(\theta, \phi)$, of varying order, l, and degree, m[1]:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} C_{lm} Y_{lm}(\theta, \phi). \tag{3.1}$$

As with a Fourier representation, we can approximate f by truncating the series to a finite number of terms. For convenience, we organize this finite collection of basis functions into a vector by the convention of encoding both order and degree with a single subscript. Thus

$$f(\theta, \phi) \approx \sum_{k=0}^{n} C_k Y_k(\theta, \phi) = \mathbf{C} \cdot \mathbf{Y}(\theta, \phi). \tag{3.2}$$

Each coefficient C_k is defined by the inner product of $f(\theta, \phi)$ with the corresponding spherical harmonic basis function:

$$\begin{aligned} C_k &= \int_0^{2\pi} \int_0^{\pi} f(\theta, \phi) Y_k(\theta, \phi) \sin\theta \, d\theta \, d\phi \\ &= \langle Y_k \,|\, f \rangle . \end{aligned} \tag{3.3}$$

This follows directly from the orthogonality of the basis functions [Wallace 84].

[1] This thesis uses the surface spherical harmonics, which form a complete orthogonal basis set for functions on the unit sphere; the more general solid spherical harmonics are defined over all space [Jones 85, pp. 1-2]. Whenever the term "spherical harmonics" is used in this thesis, the technical intent is "surface spherical harmonics".

3.2 Representing the BRDF

If we fix the incident direction (θ_i, ϕ_i), the BRDF is a function of two variables, (θ_r, ϕ_r), and the representation in Equation 3.2 suffices. To account for variation of the BRDF with incident direction, the coefficient vector $\mathbf{C}$ in Equation 3.2 can be thought of as a function of the incident direction. If a surface has isotropic roughness, as assumed in [Cabral 87] and [Sillion 91], the scattering function ρ_{bd} is independent of rotation about the surface normal, and

$$\rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r) = \rho_{bd}(\theta_i, 0, \theta_r, \phi_r - \phi_i). \tag{3.4}$$

Each coefficient C_k is a function of θ_i alone, which can be calculated for a number of selected values of θ_i and interpolated for all θ_i [Cabral 87,Sillion 91]. In general, however, a BRDF is a function of ϕ_i as well as of $(\theta_i, \theta_r, \phi_r)$, so a richer representation is needed.

3.3 Extension to Anisotropic Surfaces

For an anisotropic surface both θ_i and ϕ_i must be considered, and none of the previous representations suffices [Cabral 87,Sillion 91]. Each coefficient C_k in Equation 3.2 is thus a function of θ_i and ϕ_i:

$$\rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r) \approx \sum_{k=0}^{n} C_k(\theta_i, \phi_i) Y_k(\theta_r, \phi_r). \tag{3.5}$$

Each coefficient function, $C_k(\theta_i, \phi_i)$, is defined by the inner product of $\rho_{bd}(\theta_i, \phi_i, \cdot, \cdot)$ with the corresponding spherical harmonic basis function:

$$C_k(\theta_i, \phi_i) = \langle \rho_{bd} \mid Y_k \rangle_{\mathit{refl}} \tag{3.6}$$

where the subscript *"refl"* denotes integration over the reflection hemisphere. Reciprocity (Section 2.3.3) makes the dependence of ρ_{bd} on (θ_i, ϕ_i) exactly like its dependence on (θ_r, ϕ_r). Since spherical harmonics concisely represent the latter dependence, we also use them to represent the dependence on (θ_i, ϕ_i), expressing each coefficient function in terms of spherical harmonics. Each element of our vector $\mathbf{C}$ of coefficients is now represented in turn by a vector of coefficients, yielding a matrix $\mathbf{M}$ to represent the BRDF. Each element of the matrix $\mathbf{M}$ is given by

$$m_{jk} = \left\langle Y_j \mid \langle \rho_{bd} \mid Y_k \rangle_{\mathit{refl}} \right\rangle_{\mathit{in}} \tag{3.7}$$

where the subscripts *"in"* and *"refl"* denote integration over the incident and reflection hemispheres, respectively. Evaluation of the BRDF becomes

$$\begin{aligned} \rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r) &\approx \sum_{j=0}^{N} \sum_{k=0}^{N} Y_j(\theta_i, \phi_i) m_{jk} Y_k(\theta_r, \phi_r) \\ &= \mathbf{Y}^T(\theta_i, \phi_i) \mathbf{M} \mathbf{Y}(\theta_r, \phi_r), \end{aligned} \tag{3.8}$$

where $\mathbf{Y}(\theta, \phi)$ is the column vector of basis functions evaluated at (θ, ϕ).

The matrix representation for the BRDF can also be used to extend the global illumination method of Sillion et al. [Sillion 91], which stores reflected intensity distributions at many sample points in the environment. Since these distributions are represented by vectors of spherical harmonics coefficients, it is easy to adapt Sillion's method to use the new representation. His method shoots energy from one patch at a time, updating the distribution at each sample point for each shot. He uses spline interpolation in one dimension to get a vector $\mathbf{C}_{shot}$ of coefficients to represent the contribution of the shot. To use the new matrix representation instead, we evaluate the vector $\mathbf{Y}$ in the incident direction and multiply it by the matrix M to get the vector $\mathbf{C}_{shot}$.

$$\mathbf{C}_{shot} = \mathbf{Y}(U)\mathbf{M} \tag{3.9}$$

No other changes are needed.

3.3.1 Reciprocity

Reciprocity (Section 2.3.3) is an important physical constraint on the BRDF. As Equation 2.6 states,

$$\rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r) = \rho_{bd}(\theta_r, \phi_r, \theta_i, \phi_i)$$

for all angles of incidence and reflection. If the matrix $\mathbf{M}$ is symmetric, then

$$\mathbf{Y}^T(\theta_i, \phi_i)\mathbf{M}\mathbf{Y}(\theta_r, \phi_r) = \mathbf{Y}^T(\theta_r, \phi_r)\mathbf{M}\mathbf{Y}(\theta_i, \phi_i) \tag{3.10}$$

and the approximation in Equation 3.8 satisfies Equation 2.6. By assuring that we compute a symmetric matrix $\mathbf{M}$, we can enforce exact reciprocity; previous approaches [Cabral 87,Kajiya 85,Poulin 90,Sillion 91] afforded, at best, approximate reciprocity.

3.3.2 Storage Reduction and Filtering

The matrix **M** can be large; tens of thousands of elements are typical. Since our BRDF representation, like that of [Sillion 91], is based on spherical harmonics, we can adapt two techniques from that work to reduce the number of coefficients (and corresponding basis functions) needed: the first technique causes half the coefficients to vanish, and the second reduces the high-frequency content of the BRDF, reducing the number of coefficients needed to achieve an acceptably accurate approximation. Since we deal only with scattering to one hemisphere, we can complete the other hemisphere with an arbitrary function. We chose a function that reduces the size of the representation: $\rho_{bd}(\theta_i, \phi_i, \pi - \theta_r, \phi_r) = -\rho_{bd}(\theta_i, \phi_i, \theta_r, \phi_r)$; this causes half of the coefficients (those with $l + m$ even in the real form of spherical harmonics) to be zero; they can be omitted from the representation, reducing the matrix size by three-quarters. To economize further, we represent $\rho_{bd} \cos\theta_i \cos\theta_r$ instead of ρ_{bd}; multiplication by $\cos\theta_i$, together with the completion described above, forces C^1 continuity at the equator and drastically reduces ringing. To maintain symmetry of the matrix **M**, we also multiply by $\cos\theta_r$. Representing $\rho_{bd} \cos\theta_i \cos\theta_r$ assures that Equation 2.6 is still satisfied. This implementation detail is omitted from the following discussion.

As with a Fourier representation of a function, simply truncating all coefficients with index $l > l_{max}$ will cause ringing in the approximation, called the Gibbs phenomenon. To reduce this, we attenuate higher frequencies, as did Cabral et al. [Cabral 87], by progressively reducing the magnitude of coefficients with $l_{filter} < l \leq l_{max}$, where l_{filter} is an empirically-determined threshold. The magnitude is reduced according to a half-Gaussian with empirically-determined width.

Chapter 4

Monte Carlo Estimation of the Coefficient Matrix

> The name *Monte Carlo* was applied to a class of mathematical methods first by scientists working on the development of nuclear weapons in Los Alamos in the 1940s. The essence of the method is the invention of games of chance whose behavior and outcome can be used to study some interesting phenomena. [Kalos 86]

In many physical simulations, we must evaluate functions that defy classical closed-form solutions. Monte Carlo methods were developed to solve such problems; the early Monte Carlo simulations of the 1940's solved particle transport problems similar to optical energy transport. Since Cook et al. introduced the method to computer graphics [Cook 84], these methods have been applied more and more [Arvo 90,Kajiya 86]. This chapter introduces fundamental concepts of Monte Carlo calculations and applies them to the problem of estimating the matrix M of Chapter 3.

4.1 The Monte Carlo Method

An integral can be estimated by a sum of random samples: this is the foundation of the Monte Carlo method. Suppose that the integrand can be expressed as the product of two functions, $f(x)$ and $g(x)$, where $f(x)$ is a probability density function, i.e.,

$$f(x) \geq 0; \int_{-\infty}^{\infty} f(x)dx = 1. \tag{4.1}$$

The integral can then be estimated by

$$\int_{-\infty}^{\infty} g(x)f(x)dx = E\left(\frac{1}{N}\sum_{i=1}^{N} g(x_i)\right) \tag{4.2}$$

where the x_i are random variables distributed according to f, and $E(Z)$ is the expected value of the random variable Z. To obtain such an estimate, one would

1. Draw the N samples $x_1 \ldots x_N$ from $f(x)$
2. Evaluate $g(x)$ at each x_i
3. Compute the arithmetic mean of the results.

The simplest case of this process is estimating the definite integral of some function $g(x)$. If the probability density $f(x)$ is uniform over the integration interval (a, b), the estimator becomes

$$\int_{a}^{b} g(x)f(x)dx = E\left(\frac{1}{N}\sum_{i=1}^{N} g(x_i)\right). \tag{4.3}$$

In more complex calculations, the functions f and g can be used to simulate some physical process. In optical scattering, $f(x)$ can simulate the probability of scattering in a particular direction and $g(x)$ can model the attenuation in that direction.

One important limitation is that a Monte Carlo calculation gives an estimate of the desired integral; this estimate will vary randomly from the true value. Why would anyone tolerate the random error introduced by Monte Carlo calculations? If there is a tractable analytical solution to the integral, one wouldn't. Where analytical calculations are inadequate, the integral can often be estimated numerically; this process is called *quadrature*. Monte Carlo methods are one class of quadrature techniques. There are two main reasons to choose a Monte Carlo approach rather than deterministic quadrature: simplicity and the rate of convergence.

A Monte Carlo calculation is often the most straightforward expression of a physical problem. Optical scattering, where the number of interactions along a path is generally not known *a priori,* is a good example of this. Monte Carlo methods generalize easily to multiple dimensions; Equation 4.2 becomes

$$\int_{\infty}^{\infty} g(\mathbf{X})f(\mathbf{X})d\mathbf{X} = E\left(\frac{1}{N}\sum_{i=1}^{N} g\left(\mathbf{X}_i\right)\right) \tag{4.4}$$

where $\mathbf{X}$ is a vector of independent random variables and $f(\mathbf{X})$ is the joint probability density function for this vector.

The law of large numbers states that, given that the variance of $g(x)$ is finite, the estimate will converge (probabilistically) to the exact answer as N increases. That is if μ is the exact answer,

$$\mathrm{P}\left(\lim_{n\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N} g\left(x_i\right)\right] = \mu\right) = 1. \tag{4.5}$$

The variance σ^2 of the estimate decreases as $\frac{1}{N}$, giving expected errors (proportional to the standard deviation σ) decreasing as $\frac{1}{\sqrt{N}}$. To show this proba-

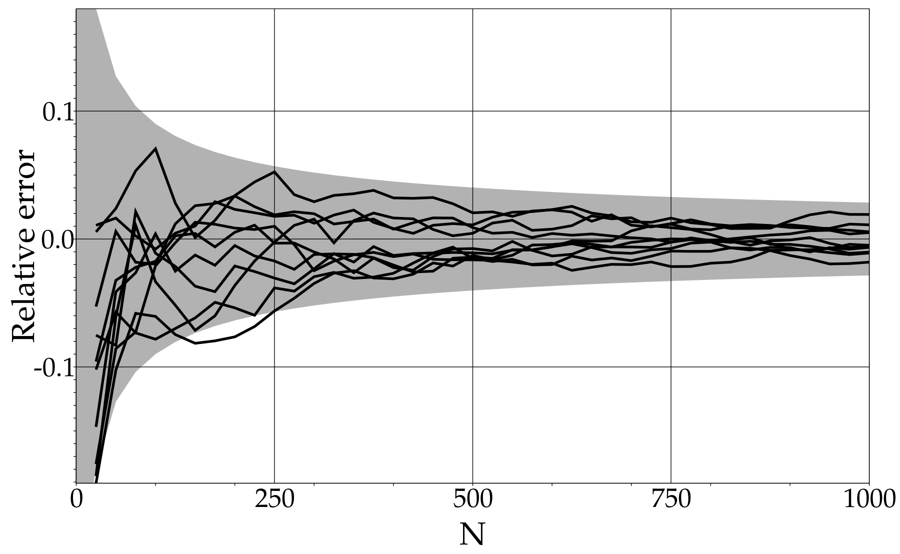

Figure 4.1: Convergence of Monte Carlo Quadrature

bilistic convergence, Monte Carlo quadrature was used to estimate the value of $\int_0^1 \cos\left(\frac{\pi}{2}x\right) dx$ as explained in [Kalos 86, pp. 94-95]. The results of ten trials of the calculation are shown in Figure 4.1. The trend of reduced error as N increases is apparent; so is the probabilistic nature of convergence. The shaded band has height proportional to $\frac{1}{\sqrt{N}}$ for reference.

Monte Carlo quadrature converges at a rate proportional to $\frac{1}{\sqrt{N}}$ regardless of dimensionality, but the convergence rate of deterministic quadrature deteriorates as the number of dimensions increases. It is always possible to choose a number of dimensions for which Monte Carlo quadrature converges faster than a deterministic method. Monte Carlo quadrature generally becomes superior to deterministic methods at levels ranging from six to ten dimensions [Kalos 86, p. 91].

4.2 Application to Estimation of the Matrix

If we bombard a specimen with incident rays from an arbitrary direction $U = (\theta_i, \phi_i)$, the BRDF can be expressed as

$$\rho_{bd}(U, V) = \frac{p(U, V)R(U, V)}{\cos\theta_r} \tag{4.6}$$

where a ray from direction U will scatter into $V = (\theta_r, \phi_r)$ with a probability density $p(U, V)$. The mean scale factor of all rays incident from direction U and scattered in direction V is $R(U, V)$.

To calculate a spherical harmonics coefficient, we must integrate the product of the BRDF ρ_{bd} and the corresponding basis function Y_k over the hemisphere.

$$\begin{aligned} C_k(U) &= \int_{S^2} \rho_{bd}(U, V) Y_k(V) dV \\ &= \int_{S^2} g_k(U, V) p(U, V) dV \end{aligned} \tag{4.7}$$

where

$$g_k(U, V) = \frac{R(U, V)}{\cos\theta_r} Y_k(V). \tag{4.8}$$

Unfortunately we have no analytical expression for p or R; we can, however, use a Monte Carlo simulation to estimate the integral in Equation 4.7. The integral can be interpreted as the expected value of $g_k(U, V)$, where V is a random variable with probability density function $p(U, V)$. If we define

$$G_k(U) = \frac{1}{N} \sum_{n=1}^{N} g_k(U, V_n) \tag{4.9}$$

where V_n are random samples distributed according to p, then the expected value of G_k is $C_k(U)$; G_k is an unbiased estimator of the integral. The rays departing from the specimen in direction V will have a mean scale factor $R(U, V)$;

this must be multiplied by $Y_k(V)/\cos\theta_r$ to give the expected value g_k for the estimator.

This leaves another integration, that with respect to U:

$$m_{jk} = \int_{S^2} C_k(U) Y_j(U) dU. \tag{4.10}$$

This integration can also be handled via Monte Carlo techniques, this time as quadrature, a discrete approximation to an integral. This is handled similarly, with the estimator

$$\widetilde{m}_{jk} = \frac{1}{N}\sum_{n=1}^{N} C_j(U_n) Y_j(U_n) \tag{4.11}$$

where the U_n are uniformly distributed over the incident hemisphere. These two sampling processes, each approximating an integral in two dimensions, can be combined into one process to approximate the four-dimensional integral desired.

$$\widetilde{m}_{jk} = \frac{1}{N}\sum_{n=1}^{N} g_k(U_n, V_n) Y_j(U_n) \tag{4.12}$$

where the U_n are distributed uniformly and the V_n are distributed according to p. $\widetilde{m}_{jk}$ is an unbiased estimator of m_{jk}.

The simulation yields $\widetilde{\mathbf{M}}$, an approximation to the symmetric matrix $\mathbf{M}$. Reciprocity of the BRDF is guaranteed only in the limit where $N = \infty$. To guarantee symmetry and improve the quality of the estimate, we average the upper triangle and the lower triangle of $\widetilde{\mathbf{M}}$ to obtain a symmetric matrix $\frac{1}{2}(\widetilde{\mathbf{M}} + \widetilde{\mathbf{M}}^T)$. The BRDF ρ_{bd} is evaluated using this matrix. The two triangles are independent unbiased estimates of the BRDF; by averaging them to obtain a symmetric matrix, we also reduce the variance of our estimate of $\mathbf{M}$. In our test cases, the approximate matrix $\widetilde{\mathbf{M}}$ is nearly symmetric, and becomes more so as N increases.

$\mathbf{M}$ is thus a sum of matrices of rank 1, each of which is the outer (matrix) product of two vectors.

4.3 The Probability Density Function p and the Scaling Function R

The probability density function $p(U, V)$ and the scaling function $R(U, V)$ are vital to the correctness of the techniques described above. They are determined by three things: the local scattering function $\widehat{p}(U, V)$, the local scale function $\widehat{R}(U, V)$, and the microgeometry. $\widehat{R}(U, V)$ is dimensionless; it serves to convert an incident energy flux to a reflected energy flux. The probability density function $\widehat{p}(U, V)$ has units of inverse steradians, since it accounts for solid angles of scattering. Any choice of the functions $\widehat{R}(U, V)$ and $\widehat{p}(U, V)$ should be checked against the following constraints:

$$\int_{S^2} \widehat{p}(U, V)\widehat{R}(U, V)dV = \int_{S^2} \widehat{\rho}_{bd}(U, V)dV \leq 1 \tag{4.13}$$

bounds energy conservation, and $\widehat{p}(U, V)$ is a probability density function in V:

$$\widehat{p}(U, V) \leq 1, \text{and} \int_{S^2} \widehat{p}(U, V)dV = 1. \tag{4.14}$$

A look at several examples of these local functions demonstrates how they are related.

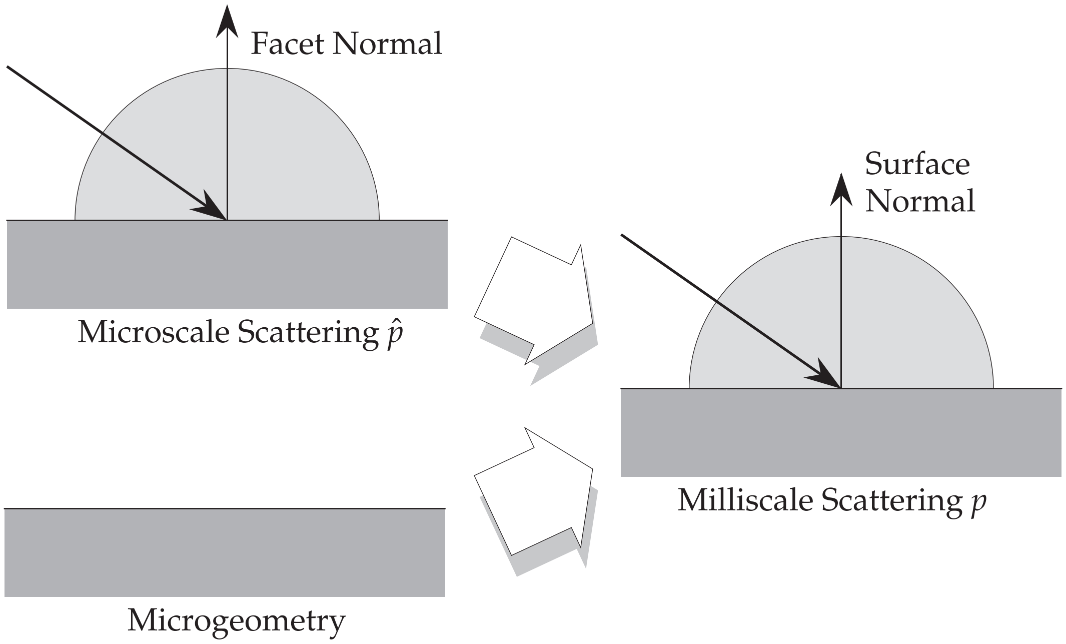

Figure 4.2: Scattering Probability of Ideal-Diffuse Flat Plate

4.3.1 Ideal Diffuse Flat Plate

In our first example, the microgeometry is an ideal-diffuse flat plate. Since there is only one planar microfacet coincident with the mean surface, the local and global scattering functions are identical: $\widehat{p} = p$, $\widehat{R} = R$, and $\widehat{\rho}_{bd} = \rho_{bd}$. We make the plate a perfect scatterer that reflects all the incident energy:

$$\int_{S^2} \rho_{bd} cos\theta_r d\omega = 1. \tag{4.15}$$

Since ρ_{bd} is constant for an ideal-diffuse surface,

$$\begin{aligned} \rho_{bd} &= \frac{1}{\int_{S^2} \cos\theta_r d\omega} \\ &= \frac{1}{\pi}. \end{aligned} \tag{4.16}$$

We can now form our estimator,

$$\rho_{bd}(U,V) = \frac{p(U,V)R(U,V)}{\cos\theta_r} \tag{4.17}$$

If we sample the reflection hemisphere uniformly, $p(U,V) = \frac{1}{2\pi}$, and

$$\begin{aligned} \widehat{R}(U,V) = R(U,V) &= \frac{\rho_{bd}(U,V)\cos\theta_r}{p(U,V)} \\ &= \frac{\frac{1}{\pi}\cos\theta_r}{\frac{1}{2\pi}} \\ &= 2\cos\theta_r. \end{aligned} \tag{4.18}$$

More generally,

$$\begin{aligned} \widehat{R} &= 2\pi\widehat{\rho}_{bd}\cos\theta_r \\ &= \left(\int_{S^2} d\omega_r\right)\widehat{\rho}_{bd}\cos\theta_r. \end{aligned} \tag{4.19}$$

$\widehat{\rho}_{bd}$ gives reflected radiance. The term 2π multiplies by the solid angle of the reflection hemisphere to convert to energy flux density, and the $\cos\theta_r$ term projects the facet differential surface area dA in the reflected direction. This is the formulation used in all the work for this thesis. Alternatively, we could distribute reflected rays according to $\widehat{p}(U,V) = \frac{\cos\theta_r}{\pi}$; then $\widehat{R} = \pi\widehat{\rho}_{bd}$ would give the correct estimator. Such a distribution would cause each reflected ray to carry a specific energy, rather than energy flux density.

4.3.2 Ideal Specular Microfacets

If the microfacets are pure ideal specular reflectors, the local scattering function $\widehat{p}(U,V)$ is a delta function; rays can scatter only in the specular direction.

$$\widehat{p}(U,V) = \delta(U,V)$$

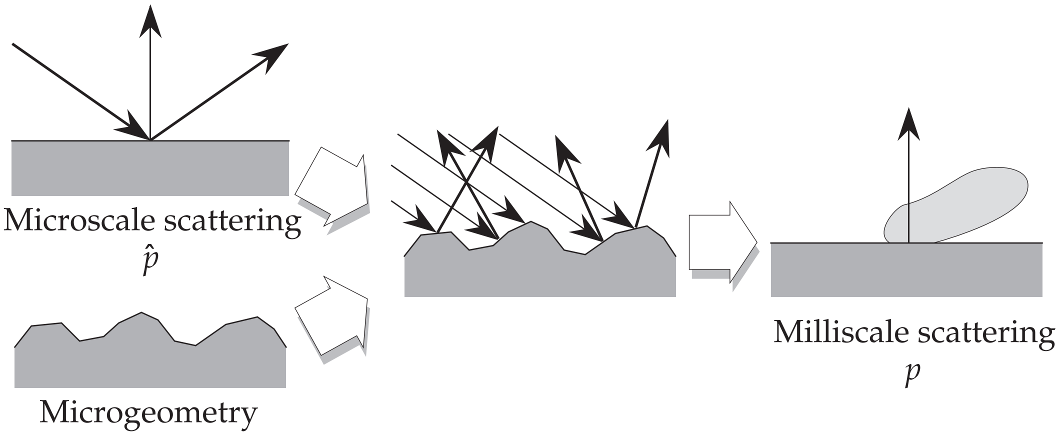

Figure 4.3: Scattering Probability of Ideal Specular Microfacets

$$= \begin{cases} \frac{1}{d\omega_i} & \text{if } V \text{ inside } d\omega_r \\ 0 & \text{otherwise} \end{cases} \tag{4.20}$$

where $d\omega_i$ is the solid angle of the incident beam and $d\omega_r$ is the solid angle of the reflected beam. The function $\widehat{p}$ remains a probability density function:

$$\int \widehat{p}(U, V) d\omega = 1. \tag{4.21}$$

The global function $p(U, V)$ is determined solely by the geometric arrangement of microfacets. The local scale $\widehat{R}(U, V)$ is the Fresnel reflectance. The Fresnel reflectance is already dimensionless, converting energy flux to energy flux; differential projected area $dA_r = dA_i$ since $\theta_r = \theta_i$. Curved microfacets will alter the differential solid angle $d\omega$, but this is is accounted for in the distribution of samples over the microgeometry.

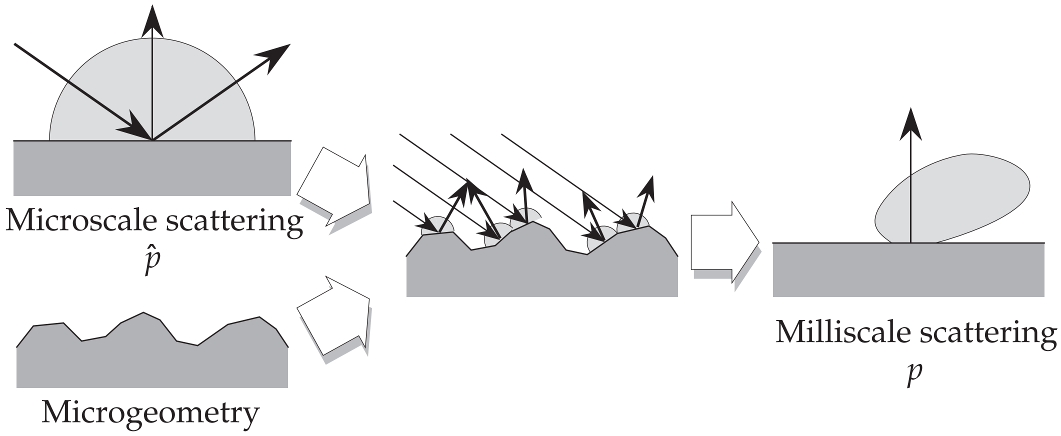

Figure 4.4: Scattering Probability in the General Case

4.3.3 General Case

The most interesting case involves both a complex microgeometry and a complex microscale BRDF. The global scattering probability p then comes from the interaction of microgeometry and microscale BRDF, and cannot generally be predicted analytically. The power of the new method lies in its ability to simulate both the global probability density function p and the mean scale factor R, even in cases where it is impossible to calculate them, or their product, analytically.

4.4 Variance Reduction

The expected error in a Monte Carlo calculation is proportional to $\frac{1}{\sqrt{N}}$, where N is the number of samples. For example, to halve the expected error of a calculation with N samples, we would have to draw $4N$ samples. Plainly any alternate method of reducing variance can offer big gains in efficiency; reducing error

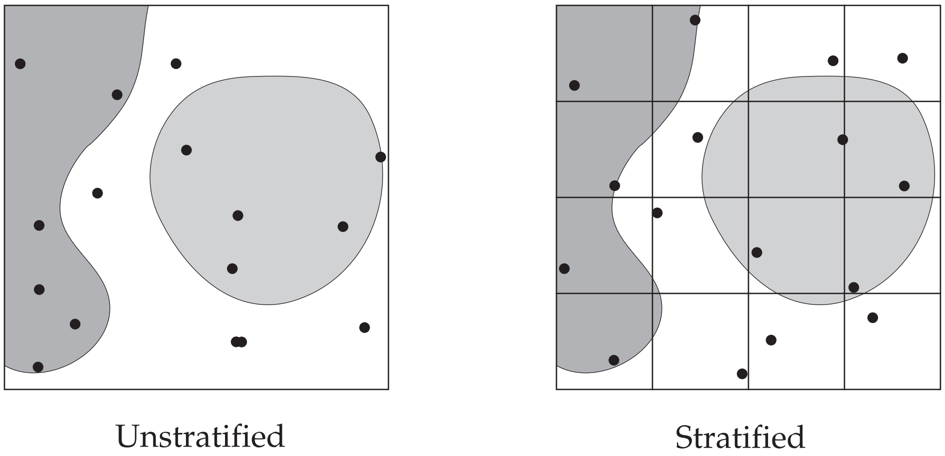

Figure 4.5: Stratified Sampling

by half effectively makes the calculation four times as fast. Most of the effort in designing good Monte Carlo algorithms goes into reducing variance. There are several major methods used to reduce variance, two of which pertain to this thesis: stratification and importance sampling.

Figure 4.5 shows a domain sampled twice with the same 16 pseudorandom number pairs. On the left, the points are distributed randomly over the entire domain; on the right, it is divided into 16 subdomains and one point is placed randomly within each. The pattern on the right is called *stratified sampling.* If the mean variance of each of the subdomains is less than the variance of the entire domain, which is likely, stratification will reduce variance. Jitter sampling in rendering [Cook 84] is a familiar example of this. Dividing the pixel area into k equal-sized subdomains and choosing a random sample point within each gives lower variance than sampling the entire area with k uniformly-distributed samples.

Another way to reduce variance is to choose a better probability distribution for the samples; this is called *importance sampling*. It replaces the "natural" probability density function $f(x)$ in Equation 4.2 with another function, $\bar{f}(x)$. The function $g(x)$ must then be replaced with

$$\bar{g}(x) = \frac{g(x)f(x)}{\bar{f}(x)} \tag{4.22}$$

so the estimator becomes

$$\begin{aligned} G &= \frac{1}{N}\sum_{i=1}^{N}\frac{g(x_i)f(x_i)}{\bar{f}(x_i)} \\ &\approx \int_a^b \left[\frac{g(x)f(x)}{\bar{f}(x)}\right]\bar{f}(x)dx \end{aligned} \tag{4.23}$$

which leaves the analytical integral unchanged, but revises the estimator.

How should the density function $\bar{f}(x)$ be chosen? The optimal probability distribution function, if $g(x)$ is non-negative everywhere in the integration domain, is

$$\bar{f}(x) = \frac{g(x)f(x)}{\int_a^b g(x)f(x)dx} \tag{4.24}$$

This distribution gives an estimator with zero variance. To use it, however, we must know the value of the integral already and the Monte Carlo calculation would be superfluous.

To reduce variance, more samples should be cast in areas where the contribution to the integral is great or where the variation of the integrand is greatest. Figure 4.6 shows the integrand used in Section 4.1, with the importance sampling function $\bar{f}(x) = \frac{3}{2}\left(1 - x^2\right)$ suggested by Kalos. This function matches the shape of the integrand, $g(x) = \cos\left(\frac{\pi x}{2}\right)$, quite closely, reducing variance by causing more

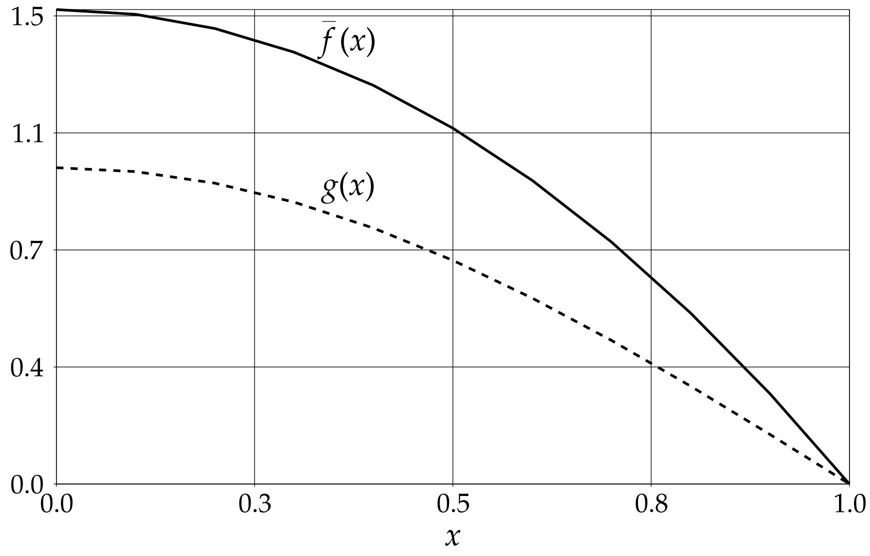

Figure 4.6: An Example of Importance Sampling

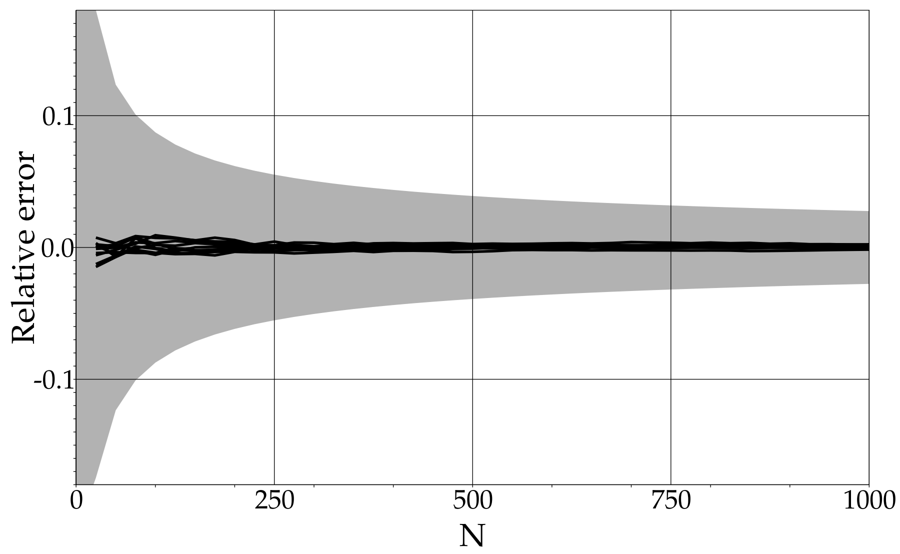

Figure 4.7: Convergence With Importance Sampling

samples to be drawn close to the left side of the interval. The magnitudes of $\bar{f}(x)$ and $g(x)$ are different because $\bar{f}(x)$ must integrate to 1 to be a probability density function. Figure 4.7, plotted at the same scale used in Figure 4.1, shows the convergence of this revised estimator. The variance has been reduced by nearly two orders of magnitude by importance sampling.

Stratified sampling is used in the work for this thesis, but importance sampling was not. See the Appendix for details.

Chapter 5

From Microscale to Milliscale

This thesis defines two levels of geometric scale of surface roughness: *microscale* and *milliscale.* The goal is to take a surface structure, *microgeometry,* and a BRDF at the microscale, and to distill them into a BRDF at the milliscale (see Figure 1.1). The milliscale BRDF will accurately represent the angular variations in scattering, but will ignore spatial variations across the microgeometry. Since our scattering calculation is based on geometrical optics, the microgeometry can only correctly model features larger than the wavelength of light being simulated; wave-optics effects can be included at the microscale by proper selection of the microscale BRDF.

This chapter explains how to integrate the microlevel scattering across an entire specimen to obtain the milliscale BRDF. It starts with basic BRDF definitions, describes the individual microscale scattering events, then explains how individual Monte Carlo events are incorporated into the milliscale model to obtain an aggregate BRDF.

At the microscale level, arbitrary reflection models may be employed, including ideal specular, ideal diffuse, and directional diffuse models. A microgeometry composed of planar ideal-specular surfaces, as in Section 4.3.2, is equivalent to geometric optics models based on microfacets, such as the Torrance-Sparrow model [Torrance 67].

We use ray tracing to model scattering events, as suggested by Cabral et al. [Cabral 87]. The ray tracer must be carefully designed to assure physically accurate results; it is described in the Appendix.

5.1 Incident Energy and the BRDF

We are estimating the BRDF ρ_{bd}, which is defined at a given wavelength as (Equation 2.5)

$$\rho_{bd}(\theta_r, \phi_r, \theta_i, \phi_i) = \frac{dI_r(\theta_r, \phi_r)}{dE_i(\theta_i, \phi_i)}$$

where dI_r is the reflected radiance and dE_i is the incident energy flux density, the incident energy per unit time per unit area. This equation holds at both micro- and milliscales.

Incident radiance I_i is defined as the incident energy flux density per unit projected area per unit solid angle

$$I_i \equiv \frac{dE_i}{\cos\theta_i d\omega_i}. \tag{5.1}$$

Thus

$$dE_i(\theta_i, \phi_i) = I_i(\theta_i, \phi_i)\cos\theta_i d\omega_i. \tag{5.2}$$

The factor $\cos\theta_i$ converts receiving area to projected area, accounting for the dependence of projected surface area on θ_i.

5.2 Integrating Over Milligeometry

We have described the possible microscale events of a single ray striking a point on the surface, but the BRDF is not intended to describe scattering from a single point on the surface. Instead, it characterizes average scattering behavior over some typical small area; this is why the technique of this thesis uses a specimen of some area of the surface. We must integrate over the specimen to get this average. Just as the Monte Carlo integration was extended in Chapter 4 to accomodate the two dimensions of the incident hemisphere, it can be extended further to integrate over a two-dimensional specimen surface. We can keep the incident energy flux density constant over all incident directions by keeping both the total incident flux and the receiving surface area constant. Equation 2.5 then becomes

$$\rho_{bd}(\theta_r, \phi_r, \theta_i, \phi_i) = \frac{dI_r(\theta_r, \phi_r)}{dE_i} \tag{5.3}$$

where dE_i is the (constant) incident energy flux density. We do this by shooting a constant number of rays (energy flux) and by distributing them over a constant surface area. The simplest way to do this is to select a fixed region of the surface and to distribute the samples uniformly over this region at each incident angle. The direction of each ray is determined by the incident angles (θ_i, ϕ_i) with respect to the mean surface; its origin will be calculated so that the ray will strike the

Figure 5.1: Target Area

notional plane of the surface, shown in Figure 5.1 in red, at the chosen sample point.

The surface region chosen should be

- much larger than the lateral geometric features of the surface, to assure a good statistical average of large-scale scattering;
- much larger than the *vertical thickness* of the surface geometry; and
- a subset of the total surface geometry.

The last two criteria serve to limit edge effects; we intend to model scattering from a small portion of an infinite surface, but rays near grazing incidence can strike the side of a finite model or miss the model entirely (Figure 5.2).

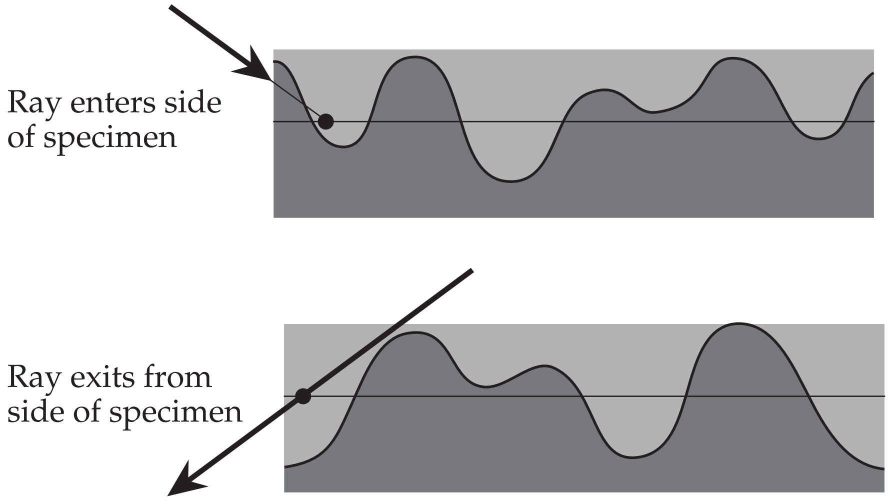

Figure 5.2: Edge effects

When a ray leaves the specimen, we update the approximate matrix $\widetilde{\mathbf{M}}$ by adding $\mathbf{Y}(U)\mathbf{Y}^T(V)R/\cos\theta_r$. The matrix, representing the BRDF ρ_{bd}, represents radiance, but each ray has an energy flux density associated with it. This flux density is converted according to

$$I_r = \frac{dE_r}{\cos\theta_r d\omega_r} \tag{5.4}$$

where θ_r is the elevation angle of reflection with respect to the notional plane. The process integrates over the entire surface visible from the reflection direction, projected onto the mean surface. Since the solid angle Since we send m rays into the sample and each represents unit incident energy flux,

$$dI_r = \frac{m}{2\pi}\frac{dE_r}{\cos\theta_r}. \tag{5.5}$$

5.3 Efficiency Considerations

We can reduce the computation needed to maintain the matrix $\widetilde{\mathbf{M}}$ by holding the incident direction U constant for several reflected directions V, updating the matrix only once for each distinct U. This happens automatically when several randomly distributed rays are spawned at each intersection, as in directional-diffuse scattering. In addition, we choose several target points on the surface for each U, further amortizing the cost of updating the matrix. (Figure 5.3). Updating the matrix then becomes a triple sum

$$\widetilde{\mathbf{M}} = \frac{1}{N_i N_p N_b} \sum_{n=1}^{N_i} \mathbf{Y}(U_n) \left\{ \sum_{m=1}^{N_p} \sum_{l=1}^{N_b} \mathbf{Y}^T(V_{ml}) \frac{R_{nml}}{\cos\theta_r} \right\} \tag{5.6}$$

where R_{nml} is the attenuation of a ray from incident direction U_n reflected in direction V_l from target point P_m on the surface. N_i is the number of incident directions used, N_p is the number of sample positions across the specimen for each incident direction, and N_b is the number of exit rays resulting from a single incident ray. This approach reduces the number of evaluations of the spherical harmonics basis functions; for $N_i N_p N_b$ samples to update the matrix, $\mathbf{Y}(U)$ is evaluated only N_i times, while $\mathbf{Y}^T(V)$ is evaluated $N_i N_p N_b$ times. The greatest savings, however, comes in matrix additions; we need only perform N_i matrix additions; the other updates simply add vectors and require far less computation.

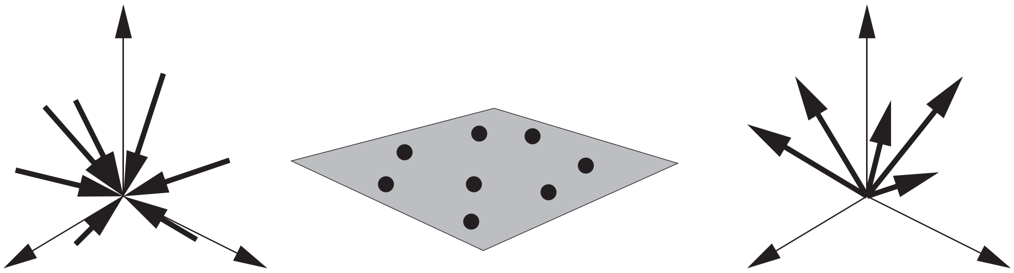

Figure 5.3: Dimensions of Integration

5.4 Convergence Measure

Since the exact matrix $\mathbf{M}$ is symmetric, we can use the asymmetry of our estimate as a measure of convergence in approximating the true BRDF. We calculate the error Q as

$$Q = \left\| \widetilde{\mathbf{M}} - \widetilde{\mathbf{M}}^T \right\| \tag{5.7}$$

where

$$\|\mathbf{A}\| = \frac{1}{N^2} \sum_{j=1}^{J} \sum_{k=1}^{J} \left| \mathbf{A}_{ij} \right| \tag{5.8}$$

where J is the size of the matrix $\mathbf{A}$. This is perhaps not as informative as a direct estimate of the variance of each coefficient, but is much cheaper to compute and tends to decline as $\frac{1}{\sqrt{N}}$, which suggests that it is directly proportional to the standard deviation of the estimate.

Chapter 6

Results

To verify and demonstrate the technique, BRDFs were generated for several surfaces textured at milliscale. Various microscale BRDFs were used, including pure ideal specular, pure ideal diffuse, and an analytical BRDF that includes wave optics effects. In one case, the method was used recursively by using a microscale BRDF which was itself generated by the new method.

All images shown in this section were generated by Monte Carlo ray tracing; the grainy texture of the images is caused by the Monte Carlo integration used to compute global illumination. Other global illumination and rendering techniques could have been used, such as that of Sillion et al. [Sillion 91].

6.1 Verification

The best means of verifying the correctness of a Monte Carlo calculation is to simulate a problem for which the solution is already known. This was done for three cases: a simple periodic surface geometry, a conical cavity, and an isotropic random surface.

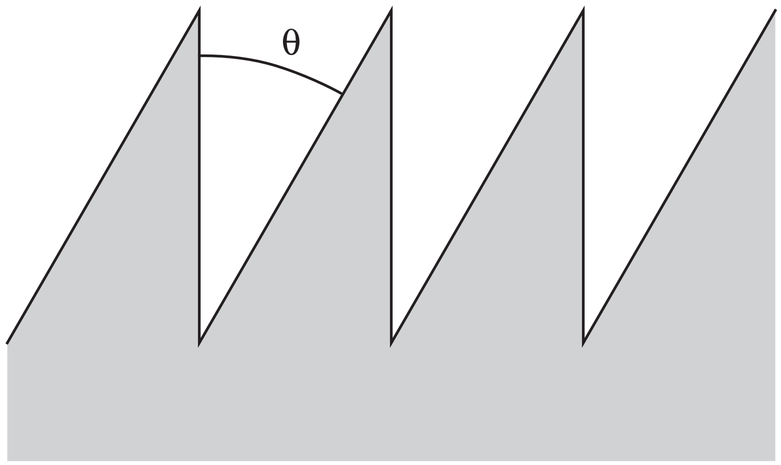

Figure 6.1: Groove Geometry (from Howell and Perlmutter)

6.1.1 Analytical ρ_{dh} for a Regular Geometry

To verify the Monte Carlo integration, the analytical work of Howell and Perlmutter [Howell 63] was duplicated. They calculated emitted energy from an asymmetric V-shaped groove (Figure 6.1) with ideal-specular reflection at the microscale. The microscale reflectivity $\widehat{\rho}_s$ in their model was constant for all incident angles $\widehat{\theta}_i$; it did not obey the Fresnel equation. The regular microscale structure of this surface gives a BRDF with severe spikes, making it difficult to approximate with smooth basis functions like the spherical harmonics. Because of this, the simulations were confined to the directional hemispherical reflectance ρ_{dh} to verify the Monte Carlo integration. The Howell paper gives emissivity ϵ as a function of angle. To calculate the directional-hemispherical reflectance ρ_{dh},

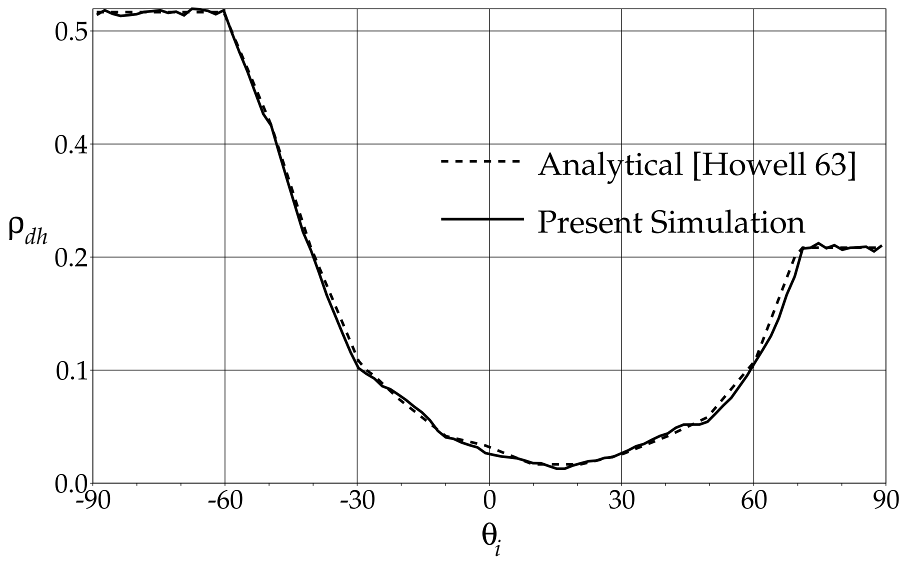

Figure 6.2: Directional Hemispherical Reflectance for Specular Grooves

we apply Kirchoff's law to obtain $\alpha = (1 - \rho_{dh})$[Siegel 81, p.73]. Two simulations were performed for a groove angle $\theta = 30°$ and a microscale reflectivity $\widehat{\rho} = 0.5$. In the first simulation, the contributions of the reflected rays were accumulated directly to estimate the function ρ_{dh}; the results are shown with a solid line in Figure 6.2, with the emittance results from Figure 4 of the Howell/Perlmutter paper, transformed to ρ_{dh}, as a dashed line. 40,000 rays were shot from each of 100 incident directions. The Howell/Perlmutter values were obtained by careful measurement from their Figure 4. The simulation agrees almost exactly with the analytical model; this shows that the Monte Carlo technique is capable of excellent accuracy in estimating ρ_{dh}.

In the second simulation, the contributions of the reflected rays were used to estimate spherical harmonics coefficients to approximate the function ρ_{dh}.

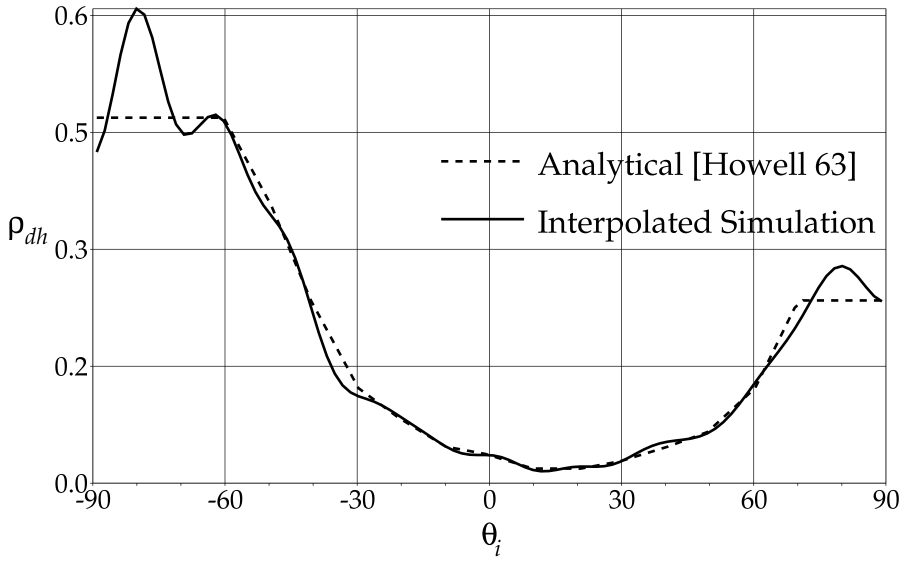

Figure 6.3: Directional Hemispherical Reflectance for Specular Grooves

Figure 6.3 shows the results of this simulation with a solid line, with the Howell/Perlmutter results again shown as a dashed line. About 1.7 million rays were shot from random incident directions to estimate the coefficients; coefficients up to $L = 23$ were used, for a total of 273 coefficients. The result shows good agreement with the Howell/Perlmutter results, but shows some ringing. The regular structure gives a ρ_{dh} with discontinuous slope, which is difficult to closely approximate with a finite set of smooth basis functions.

6.1.2 Monte Carlo Simulation for a Conical Cavity

The method and implementation were further verified by replicating some work of Polgar and Howell [Polgar 65]. Their paper described a Monte Carlo simulation of scattering from a conical cavity with an ideal diffuse BRDF at microscale.

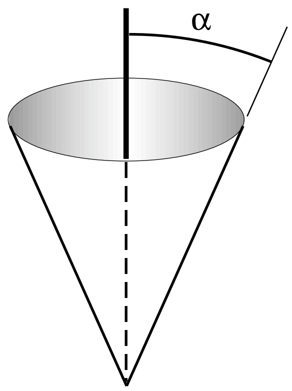

Figure 6.4: Geometry of Conical Cavity (from Polgar and Howell)

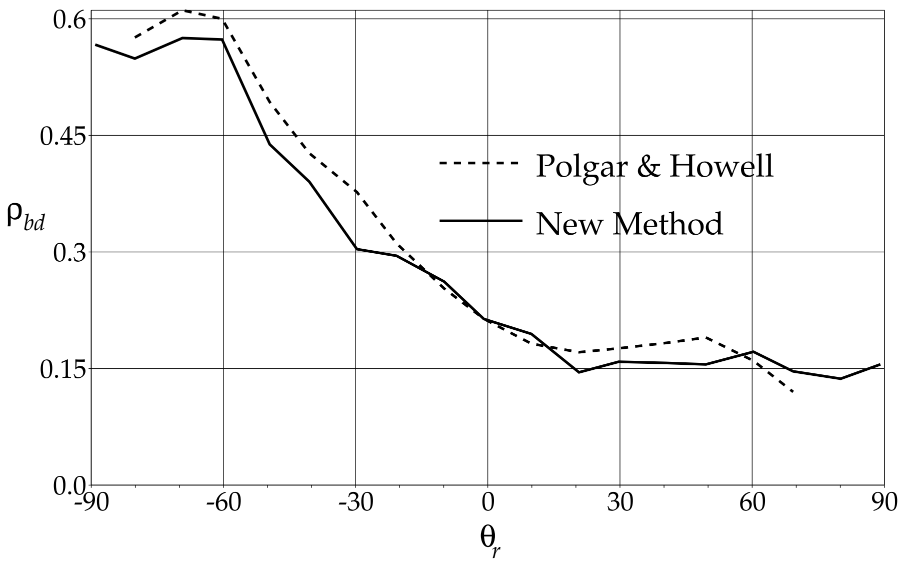

Figure 6.5: BRDF for Conical Cavity (θ_i = 60°)

One simulation was run for a cone with half-angle α of 15° (Figure 6.4). The cone was approximated as 400 triangles. Microscale reflectivity ρ_{dh} (ρ_s in Polgar and Howell's terminology) was 1. Polgar and Howell report directional reflectance (their $\rho_{\xi,\xi'}$) as energy flux density; this was measured from their Figure 8 and divided by $\cos\theta_r$ to convert to intensity for comparison. Figure 6.5 shows Polgar and Howell's results with a dashed line; the results from the new method are shown with a solid line. The agreement is good. This simulation spawned only one ray at each scattering event (see Appendix A); simulations that spawned multiple rays failed to agree with Polgar and Howell's results. This seems to be caused by bias in the ray termination procedure (see Appendix).

6.1.3 Analytical ρ_{bd} for an Isotropic Surface

The method was also compared to that of He et al. [He 91] as further confirmation of its accuracy. The He model assumes a flat Gaussian-rough surface for which, at the microscale, the surface is an ideal specular reflector; it models full wave optics effects for such a rough surface. Reflection is governed by geometric optics in the new method; shadowing and masking effects of the surface are included through the occlusion calculations in the ray tracer. Wave optics effects were not included, except for the Fresnel coefficient for each microfacet.

White noise was filtered via an FFT to create Gaussian height fields, and the resulting points were connected by triangular facets. Each facet was modeled as a perfectly smooth mirror of pure aluminum. To integrate over a specimen sufficiently large compared to the roughness height, an area of 8×8 millimeters was

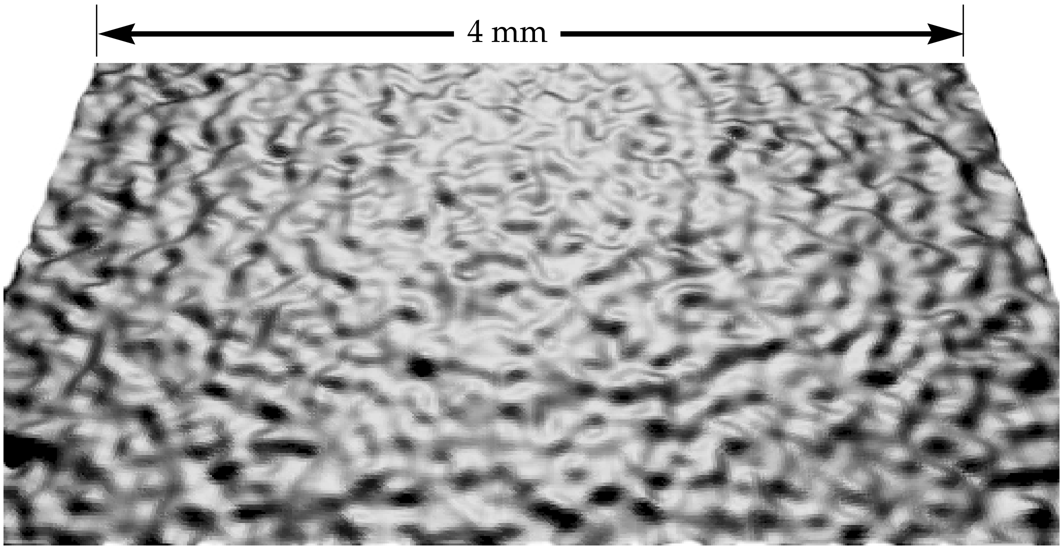

Figure 6.6: A Gaussian Surface

used. To assure adequate representation of the surface, a total of 524,288 polygons was used. The model was created in four sections of 131,072 polygons, each generated with a different random number seed, to represent a square patch of surface 4mm wide. One of the sections is shown in Figure 6.6. The roughness length parameters of the surface were $\sigma = 10\mu m$ vertically and $\tau = 65\mu m$ horizontally (Figure 6.6). The simulation used a specimen patch 3.13mm wide in the center of the geometric model; this assured that no incident ray could intersect the "side" of the patch within 2σ of the notional plane.

Incident ray angles were restricted to $\theta_i \leq 88^\circ$ to keep the effective roughness greater than 460 nm, the shortest wavelength employed. This kept behavior in the regime where geometric optics is valid; were the wavelength to approach the size of surface features, wave-related effects would begin to affect

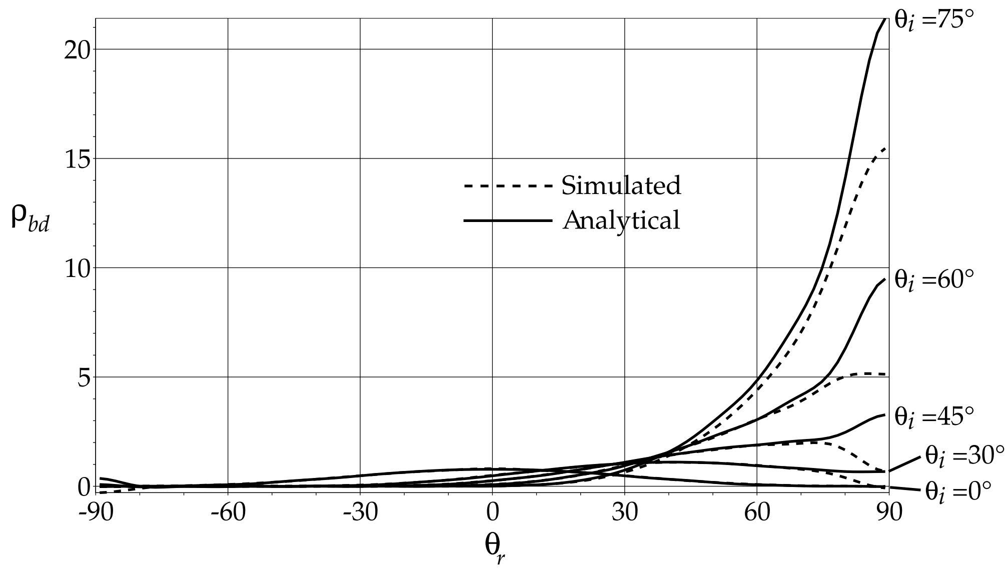

Figure 6.7: Comparison of BRDF with Model of He

the scattering. Results are plotted as solid lines in Figure 6.7 for incident angles $\theta_i = 0^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ$. Dashed lines show results from the model of He [He 91], which assumes a Gaussian rough surface and allows for wave optics effects. The He model is shown in the limit of large surface roughness, $\sigma \gg \lambda$, where wave optics effects should be negligible.

The simulation agrees well with the analytical model for reflection angles less than about 80°, but diverges for greater angles. To isolate this discrepancy, the directional hemispherical reflectance ρ_{dh} was calculated. This is plotted in Figure 6.8 for three cases: the new method using spherical harmonics to interpolate ρ_{dh}, the model of He, and raw Monte Carlo integration with no interpolation. The interpolated result, shown with a solid line, is plainly wrong for grazing incident angles: ρ_{dh} cannot physically be greater than 1 (Section 2.3.4). The raw

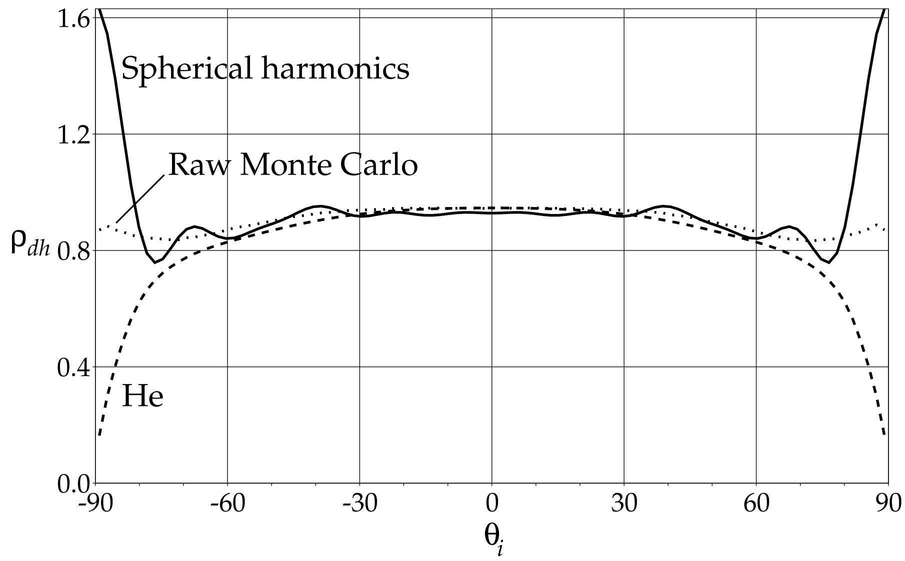

Figure 6.8: Comparison of ρ_{dh} with Model of He

Monte Carlo result, plotted with a dotted line, is physically plausible, though it does not match He's model (dashed line). The divergence at large angles could result from different treatment of shadowing and masking effects near grazing.

The interpolated result diverges from the raw Monte Carlo results near grazing incident angles; this suggests that the interpolation introduces error. The matrix approximates $\rho_{bd} \cos\theta_i \cos\theta_r$, so the result must be divided by $\cos\theta_i \cos\theta_r$ to obtain the BRDF. If the error in approximating this function is roughly constant over the hemisphere, dividing by $\cos\theta_r$ to recover ρ_{bd} will magnify the error near the horizon (i.e. as $\theta_r \rightarrow \frac{\pi}{2}$). Figure 6.8 shows just such a trend.

The error in ρ_{dh} at large incident angles is disturbing, but its effect on energy conservation is small. Recall that the BRDF ρ_{bd} gives a radiance value I_r; the energy E_r scattered in any reflected direction (θ_r, ϕ_r) is proportional to $I_r \cos\theta_r$, reducing the effect of the error at high angles of reflection.

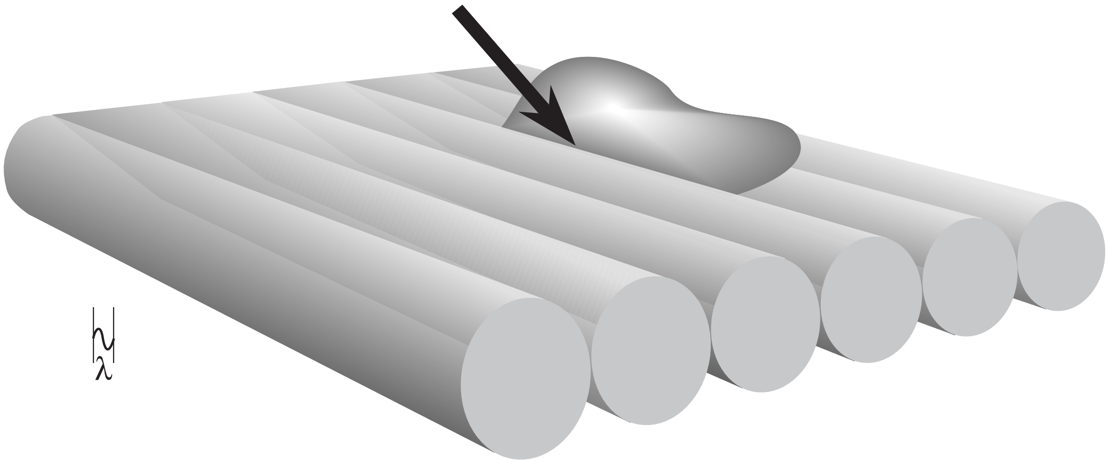

Figure 6.9: Parallel Cylinder Model of Anisotropic Surface

6.2 Simple Anisotropy

The method can create an anisotropic milliscale BRDF from an isotropic microscale BRDF; the new technique simulates larger-scale anisotropy. The geometry shown in Figure 6.9 was used with He's analytical BRDF model [He 91] to demonstrate this. Each of the parallel cylinders had a microscale BRDF calculated with He's method; this simulated slightly rough anisotropic aluminum surface at the microscale, including wave optics effects. The longest simulated wavelength $\lambda = 640nm$, as shown approximately to scale in Figure 6.9, was small compared to the size of the cylinders. Figure 6.10 shows 3-dimensional plots of the resulting BRDF. The two plots differ in that the surface has been rotated 90°. The bottom plot shows two lobes because the microgeometry resembles V-shaped grooves; each side of the grooves creates a lobe by reflecting light in a particular direction.

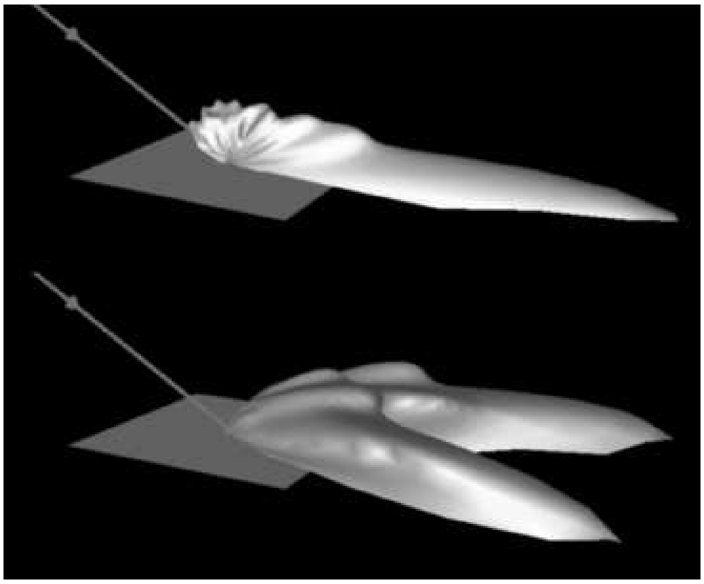

Figure 6.10: Plots of Anisotropic BRDF

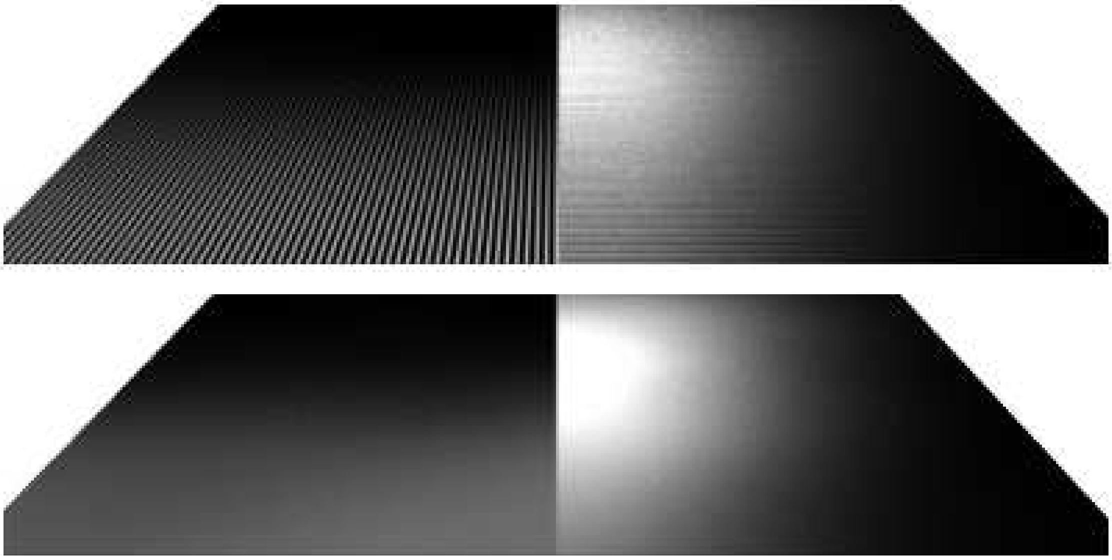

Figure 6.11: Parallel Cylinder Model vs. Anisotropic BRDF

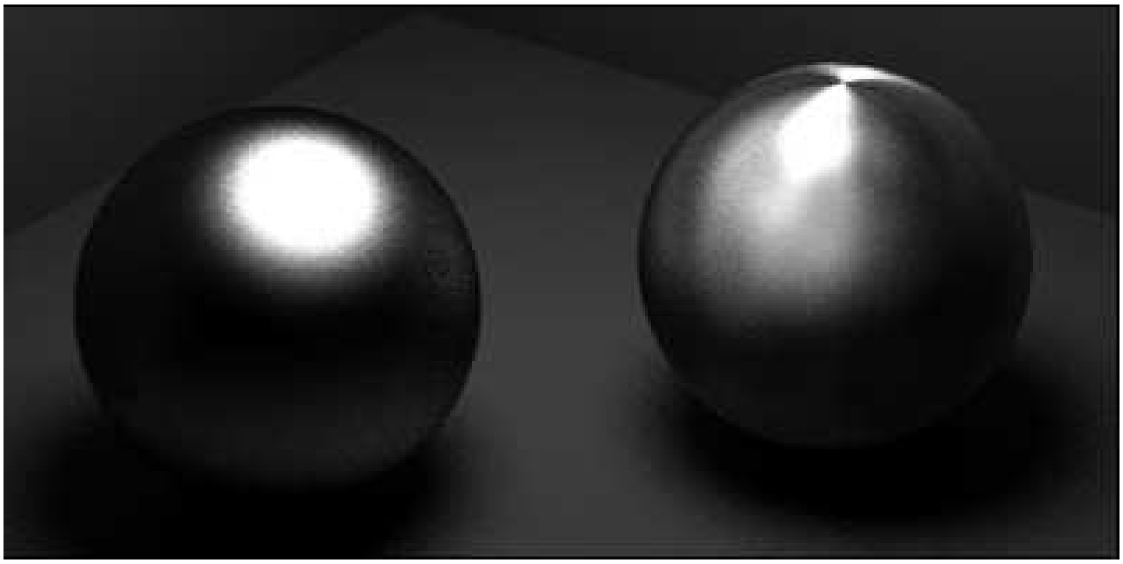

Figure 6.12: Isotropic and Anisotropic Aluminum

How well does this milliscale BRDF capture the area-averaged scattering from this microgeometry? Figure 6.11 compares a rendering of the microgeometry with a rendering using the milliscale BRDF. The top half of the figure shows the microgeometry. On the left side, the cylinders are oriented with axes perpendicular to the screen; on the right side the axes are parallel to the screen. Note how the surface orientation affects the appearance, revealing the anisotropic behavior of the surface. The bottom half of the figure shows a similar scene, but with two flat plates replacing the arrays of cylinders. Both plates use the milliscale BRDF shown in Figure 6.10. In the left half of the figure, the axis of anisotropy was oriented perpendicular to the screen; in the right half, it is oriented parallel to the screen. The scattering patterns are similar; when viewed from a distance, the images look the same. The microscale BRDF is important

Figure 6.13: Anisotropic Aluminum Teapot

for generating the upper images; the milliscale BRDF is used for the lower figures. The anisotropy is further illustrated in Figure 6.12, where the same object is rendered with two BRDFs for rough aluminum, one isotropic and one anisotropic. Figures 6.13 and 6.14 show an aluminum teapot and an aluminum automotive wheel created using this anisotropic BRDF. The polishing scratches were oriented as shown in the inset diagrams. The energy-consistency of the BRDF, not guaranteed by previous approaches [Kajiya 85,Poulin 90], allows an accurate global illumination solution.

Figure 6.14: Anisotropic Aluminum Wheel

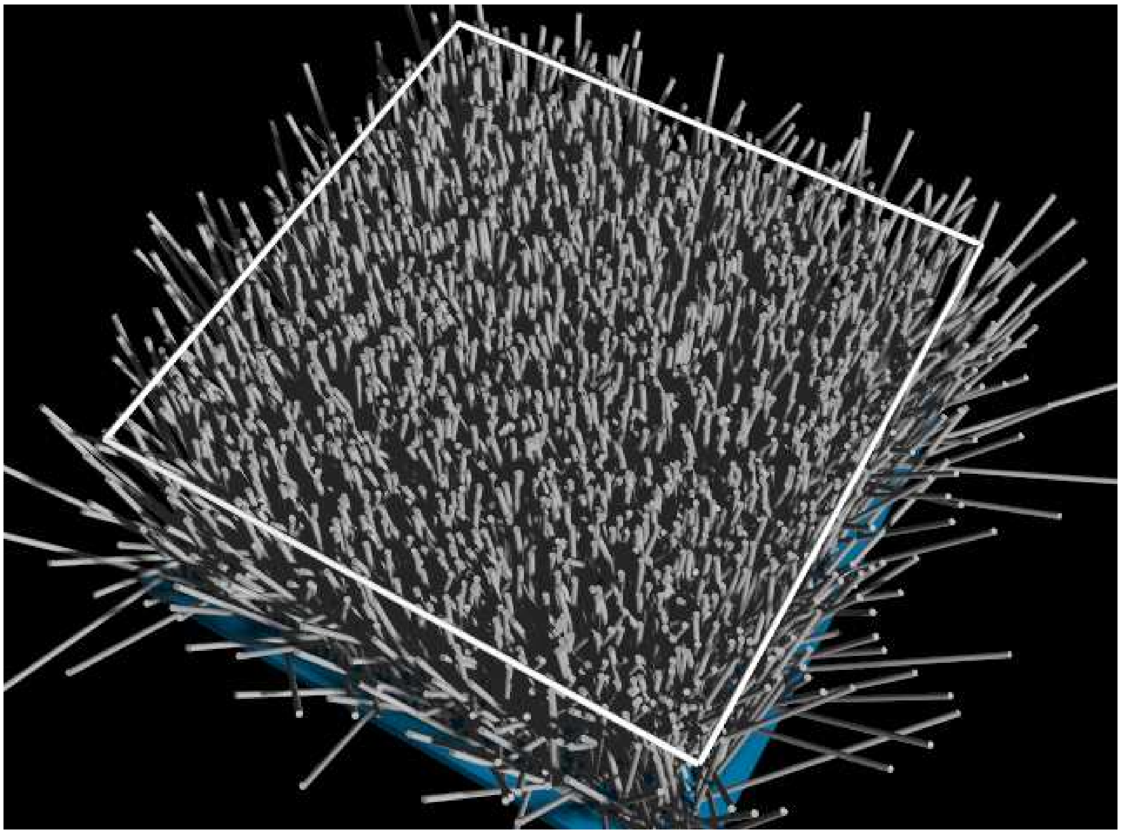

Figure 6.15: Microgeometry for Velvet

6.3 Velvet

A more complex microgeometry is that of velvet: it consists of many specular transparent fibers extending from a fabric base in roughly parallel directions. This was modeled as a forest of narrow cylinders, with the angle of each cylinder perturbed randomly (Figure 6.15). The target area for incident rays is shown at the top of the fibers. The flat ends of the cylinders would constitute a surface of microfacets and result in white directional-diffuse scattering, unlike real velvet. Since real fibers do not have ideally smooth flat ends, each fiber in the simula-

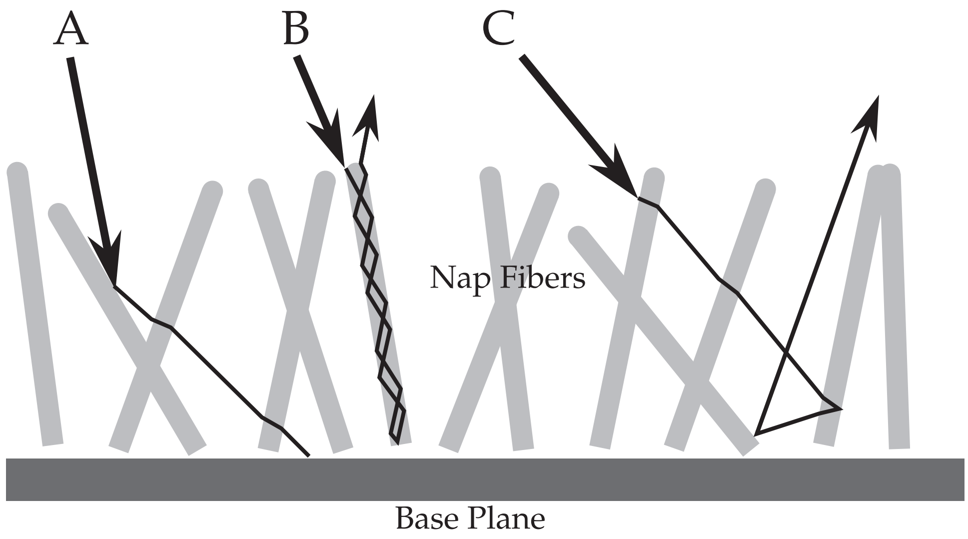

Figure 6.16: Ray Paths Through Velvet

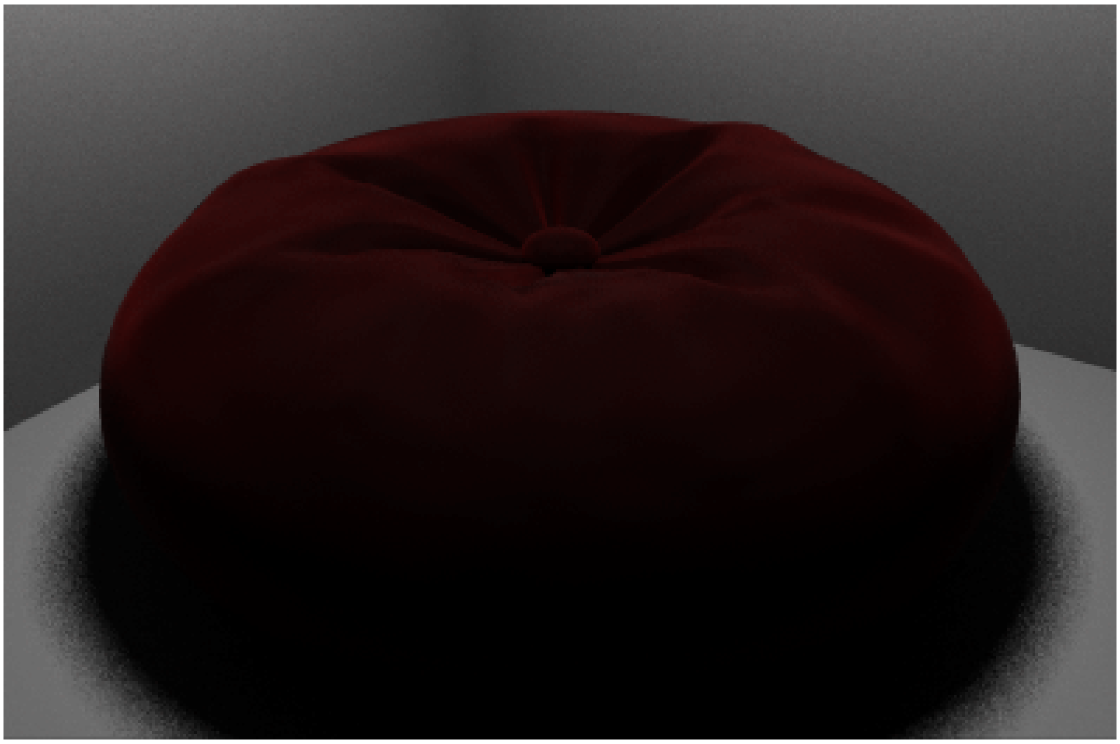

Figure 6.17: Velvet Cushion

tion was topped with a hemispherical end to approximate random rough ends. In Figure 6.15 the fibers are shown as ideal diffuse for clarity. In the BRDF simulation, the fibers were a transparent ideal-specular plastic material. The index of refraction was 1.5 in all wavelengths; extinction coefficients (see Appendix A) were $K_\lambda = (5, 45, 45)$ for the red, green, and blue channels, respectively. Whenever a ray intersected a fiber, it was either reflected (with probability equal to the Fresnel reflectivity) or transmitted; when it intersected the base plane, it was absorbed. All scattering was from the nap of the velvet, not from the base weave. Figure 6.16 shows three examples of the possible long, complex paths that a ray could follow. Ray A is refracted and reflected until it intersects the base plane and terminates, contributing nothing to the BRDF. Ray B is completely reflected internally, strikes the bottom end of the fiber, and is reflected back out. Ray C is refracted and reflected, eventually re-emerging. Both ray B and ray C contribute to the BRDF. Figure 6.17 shows an image made using the resulting BRDF.

6.4 Woven Cloth

The method can also be used recursively to model several scales of roughness; this is demonstrated by modeling woven cloth as shown in Figure 1.1. Figure 6.18 shows the microgeometry used for the weave pattern of the cloth. An anisotropic BRDF was used to model the scattering from individual fibers in the threads. The microgeometry, as with anisotropic aluminum, was a bed of parallel cylinders. The microscale BRDF, however, was a Fresnel reflectance function

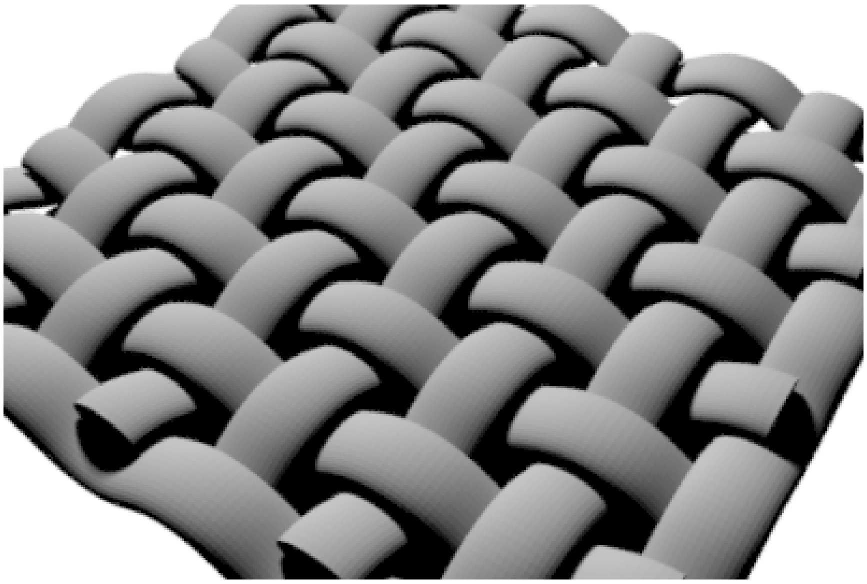

Figure 6.18: Structure of Cloth Model

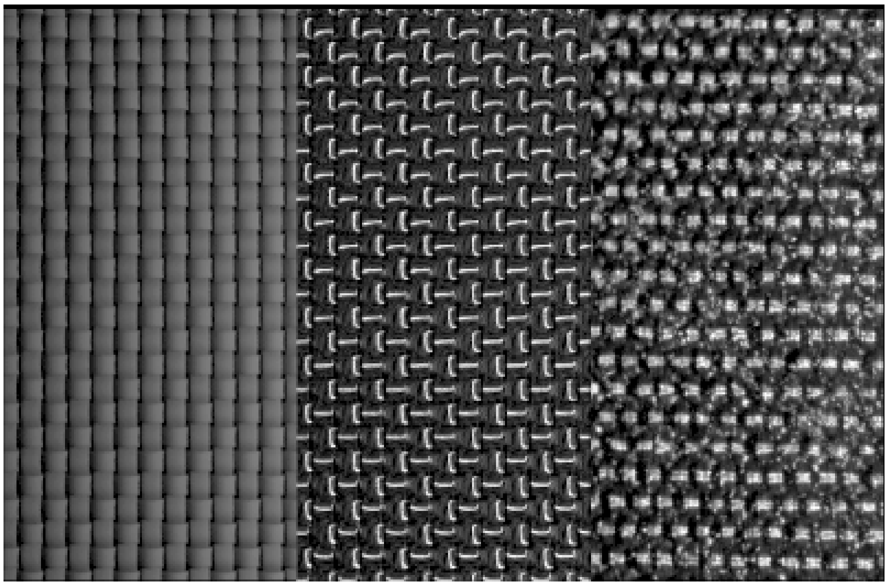

Figure 6.19: Cloth Model Compared With Real Cloth

Figure 6.20: Nylon Cushion

with index of refraction 1.5 to simulate black synthetic fibers. Figure 6.19 has three parts: on the left, the cloth microgeometry is shown with an ideal-diffuse BRDF; in the center, it is shown with the thread BRDF, and on the right is an image scanned at high magnification from a cloth sample. Figure 6.20 shows a cushion upholstered in black nylon, rendered using the BRDF obtained from this process.

6.5 Summary

The method has shown its usefulness in simulating a wide variety of surfaces, ranging from brushed aluminum to velvet. It has also shown good correlation with analytical models and another Monte Carlo simulation.

In summary, the general procedure decribed in this chapter to calculate a BRDF is as follows:

- Decide the breakdown of scales (microscale vs. milliscale)
- Create the microgeometry
- Obtain a microscale BRDF, either from an analytical model or from another simulation
- Run the simulation

The result at this point is a file containing the matrices for the BRDF at several wavelengths (3 in all the cases above). This is then used in the global illumination and rendering phases.

Chapter 7

Conclusion

This thesis introduces a very general method for obtaining a BRDF from surfaces of complex microgeometry. Not only is the method general, but its BRDF representation and careful physical simulation assure greater physical accuracy than previous numerical methods [Kajiya 85,Cabral 87]. This method does not supplant analytical models, but complements them. It can simulate complex microgeometries which defy analytical models, but relies on these models to include wave optics effects. Furthermore, the method is the first to allow recursive use to simulate surfaces with several scales of texture. A complex surface with several natural levels of scale, such as woven cloth, can be simulated in several steps, by using the milliscale BRDF resulting from one simulation as the microscale BRDF in the next. The method agrees well with both analytical models and a previous Monte Carlo calculation.

The resulting BRDFs are represented in a form that is concise and well-suited for use in rendering and global illumination calculations. The representation not only enforces exact reciprocity, but it also assures smoothness, to avoid visual artifacts like quantization and Mach bands. It can be easily evaluated, so it is directly useful in ray tracing, including stochastic global illumination methods [Kajiya 86,Ward 88]. The global illumination technique of [Sillion 91] can be easily extended to accommodate the new representation.

Generating BRDFs in this way can be very expensive; some of the BRDFs computed for over a week on a powerful workstation (Apollo DN10000VS). Once computed, however, the BRDF can be used again and again in renderings and visual simulations.

More can be done to optimize and improve the technique. First, computation could be drastically reduced in many cases by simple variance reduction techniques. Importance sampling of the microscale BRDF $\hat{\rho}_{bd}$ would be the obvious place to start.

Second, no effort has been made to exploit any structure of the matrix $\mathbf{M}$, either in the simulation itself or as a postprocess before using the BRDF. For example, if the surface shows any symmetry about the surface normal, some terms in the matrix become zero. If this is known before the simulation, the computation could be sped up by reducing the size of the matrix and reducing the number of basis functions to be evaluated for each ray. The resulting matrix from any simulation could also be examined for insignificant values. Deleting these could compress the matrix further and increase the speed of evaluation, since the major expense of evaluation is the matrix multiplication involved.

Third, more work is needed in predicting and measuring error. We have developed no method to measure how many coefficients are needed for a given BRDF. We also have no real way to measure when the variance of the Monte Carlo estimation has declined enough. This is difficult because the variance of a single coefficient does not affect the variance of the reconstructed function in any simple way.

Fourth, the BRDFs might be represented in an another manner for efficiency in evaluation or importance sampling.

Finally, many other complex surfaces may yield to this method. For example, metallic paint contains many flakes of metal suspended in a transparent medium. For remote sensing, the method could give accurate predictions of reflection from vegetation canopies [Goel 91]. It is not limited to conventional ray tracing of surfaces; given a volume-density ray tracer, such as described by Kajiya [Kajiya 84], the method could give BRDFs for clouds (e.g. for remote sensing) or texels [Kajiya 89].

Appendix A

An Energy-Consistent Ray Tracer

To simulate scattering accurately, we must account for all energy in the environment, assuring that energy is neither created nor lost. To do this, the simulation must maintain the energy balance of each scattering event. The ray tracer used in this thesis models three kinds of events, shown in Figure A.1. It associates an energy flux density with each ray; this is attenuated at each bounce (and divided into several rays for diffuse scattering), emerging as energy flux density for the next interaction. Energy flux density is converted to radiance, the proper quantity for the BRDF, when a ray finally exits the model. The radiance is integrated over the specimen, as explained in Chapter 5, to create a milliscale BRDF.

A.1 Specular Reflection

The BRDF at the microscale may contain an ideal specular component $\widehat{\rho}_s$. Whenever a ray hits such a microfacet, we model the transfer by spawning a ray in the

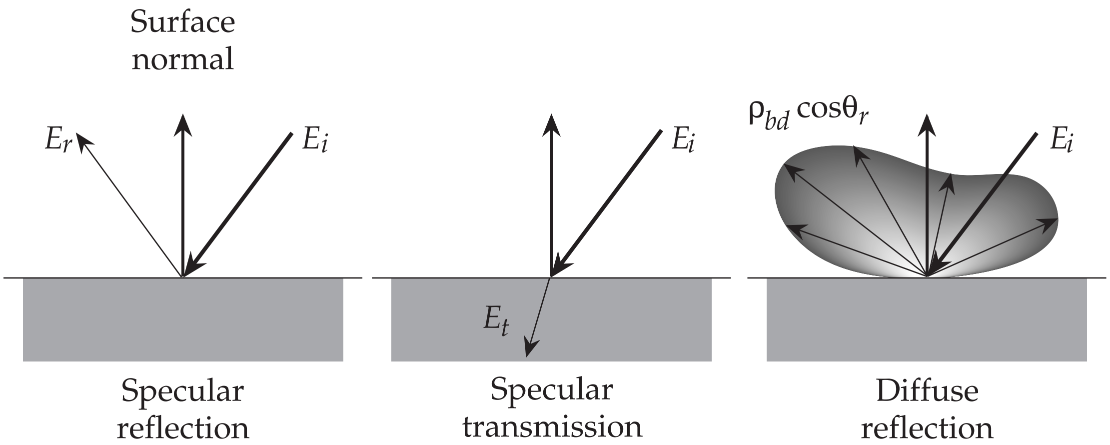

Figure A.1: Local Scattering Modes

specular direction as in classical ray-tracing [Whitted 80]. The energy flux density of this ray is

$$dE_r = \widehat{\rho}_s(\widehat{\theta}_i)dE_i \tag{A.1}$$

where dE_i is the flux density of the incident ray, $\widehat{\theta}_i$ is the incident elevation with respect to the local facet, and $\widehat{\rho}_s$ is the specular reflection coefficient for the facet. For a perfectly smooth surface, $\widehat{\rho}_s$ is the Fresnel reflectance; for rougher surfaces, $\widehat{\rho}_s$ must be calculated with a more general wave-optics model (e.g. [He 91]). If the surface is very rough, with roughness features much larger than the wavelength of light, $\widehat{\rho}_s = 0$.

The ray tracer uses *path tracing* [Kajiya 86] to reduce the number of rays cast. Instead of casting two rays from every ideal-specular interaction, one for reflection and one for subsurface interaction, it randomly sends a reflection ray with probability $\widehat{\rho}_s$.

$$dE_r = \begin{cases} dE_i & \text{if } \xi < \widehat{\rho}_s(\widehat{\theta}_i) \\ 0 & \text{otherwise} \end{cases} \tag{A.2}$$

where ξ is a random number from a uniform distribution over $(0, 1)$. If no reflection ray is sent, a subsurface interaction (either directional-diffuse reflection or specular transmission) is allowed to take place. This gives a stochastic approximation of the reflectance:

$$\langle dE_r \rangle = \widehat{\rho}_s(\widehat{\theta}_i) dE_i. \tag{A.3}$$

A.2 Specular Transmission

The method may be used to model microgeometries that include transparent materials. Whenever a ray encounters a smooth interface between media of different refractive indices, we must calculate the energy transfer through the interface. Neither energy flux density nor radiance is preserved at the interface [Hecht 87, p. 102], since solid angles are altered. Hecht gives the determining relationship as

$$\begin{aligned} AI_t \cos\widehat{\theta}_t &= AI_i \cos\widehat{\theta}_i - AI_r \cos\widehat{\theta}_r \qquad \text{(A.4)} \\ &= A \cos\widehat{\theta}_i \,(I_i - I_r) \end{aligned}$$

where I_i, I_r, and I_t are the incident, reflected, and transmitted intensities, respectively, and A is the area covered by the incident beam. A similar change in radiance occurs when the light exits from the transparent medium. This effect is often ignored in ray tracing; in many cases, like transmission through a thin plate or through a sphere, the changes in radiance and energy flux density cancel because the angle of transmission from the far surface is equal to the initial angle of incidence. Sometimes, as in transmission through one end and one side of a cylinder, this is not true and we must model the effect. In the Monte Carlo simulations described in this thesis, the change in solid angle is accounted for by the distribution of rays. Rays within some incident solid angle $d\omega_i$ will be refracted and scattered through the correct transmitted solid angle $d\omega_t$. Path tracing, described in Section A.1, simulates the attenuation of the transmitted ray stochastically:

$$\begin{aligned} dE_t &= \widehat{\tau}_s(\widehat{\theta}_i) dE_i \\ &= \left[1 - \widehat{\rho}(\widehat{\theta}_i)\right] dE_i. \end{aligned} \tag{A.5}$$

For this thesis, only transmission through perfectly smooth surfaces is modeled, so the reflectivity $\widehat{\rho}$ is the Fresnel reflectivity $\widehat{\rho}_s(\widehat{\theta}_i)$.

We must also model the attenuation of the ray as it passes through an imperfectly transparent medium, such as colored glass. This is governed by Bouguer's law [Siegel 81, pp. 424-427]:

$$I_\lambda(S) = I_\lambda(0) e^{\int_0^S K_\lambda(S) dS}. \tag{A.6}$$

where S is the path distance through the medium ($S = 0$ at the entry point) and K_λ is an extinction coefficient determined by the material and wavelength. For a uniform medium, $\int_0^S K_\lambda(S) dS = K_\lambda S$ and the ray is attenuated by

$$I_\lambda(S) = I_\lambda(0) e^{K_\lambda S}. \tag{A.7}$$

Since solid angles are not altered, Equation A.7 is also valid when intensities are replaced by energy flux densities.

A.3 Directional-Diffuse Reflection

The most complex transfer takes place when a ray strikes a facet that shows directional-diffuse scattering. When a ray hits one of these facets, we send out several rays in random directions within the reflection hemisphere. This approximates scattering according to the ideal-diffuse and directional-diffuse parts of the BRDF. The total energy transfer is determined by

$$\begin{aligned} dE_r(\widehat{\theta}_r, \phi_r) &= dI_r dA d\omega_r \cos\widehat{\theta}_r \\ &= dE_i \widehat{\rho}_{bd}(\widehat{\theta}_i, \phi_i, \widehat{\theta}_r, \phi_r) dA d\omega_r \cos\widehat{\theta}_r \end{aligned} \tag{A.8}$$

where dE_r is the reflected energy flux density in a particular direction, $\widehat{\rho}_{bd}$ is the diffuse part (including directional-diffuse) of the microscale BRDF, dA is the differential area on the surface, and $d\omega_r$ is the solid angle of reflection. $(\widehat{\theta}_r, \widehat{\phi}_r)$ is the reflection direction with respect to the local facet. We multiply by $d\omega_r \cos\widehat{\theta}_r$ to convert the radiance given by $\widehat{\rho}_{bd}$ to energy flux density for the next scattering event.

In our implementation the reflected rays are uniformly distributed in solid angle, so

$$d\omega_r = \frac{2\pi}{n} \tag{A.9}$$

where 2π is the total solid angle of the hemisphere and n is the number of reflected rays shot.

A.4 Selection of the Scattering Probability Density $\widehat{p}$

The ray tracer described here uses a uniform local scattering probability density $\widehat{p}(U, V) = \frac{1}{2\pi}$; this is by no means the only choice or the most desirable. Many scattering distributions will yield an unbiased estimator of $\widehat{\rho}_{bd}$, provided the functions $\widehat{p}$ and $\widehat{R}$ are chosen as a consistent pair; but they may result in widely different variances.

Uniform scattering probability lies at one end of a range of alternatives that include scattering with probability density proportional to $\cos\widehat{\theta}_r$ and scattering proportional to $\widehat{\rho}_{bd}\cos\widehat{\theta}_r$.

Distributing rays according to $\widehat{\rho}_{bd}\cos\widehat{\theta}_r$ is an example of importance sampling (Section4.4), since samples would be concentrated in areas with most contribution to the integral. Such a distribution would embody the entire scattering function in $\widehat{p}$; $\widehat{R}$ would be constant for all V. This assumes that $\widehat{\rho}_{bd}\cos\widehat{\theta}_r$ can be sampled directly; this is straightforward for some BRDF approximations, such as a Gaussian distribution [Ward 92] or the Phong model [Shirley92], but the approximation used for this thesis (Chapter 3) cannot efficiently be sampled directly.

For ideal-diffuse surfaces, $\widehat{\rho}_{bd}$ is constant, so $\widehat{p} = \frac{\cos\widehat{\theta}_r}{\pi}$ distributes rays according to $\widehat{\rho}_{bd}\cos\widehat{\theta}_r$ and would reduce variance; it would send more rays toward the

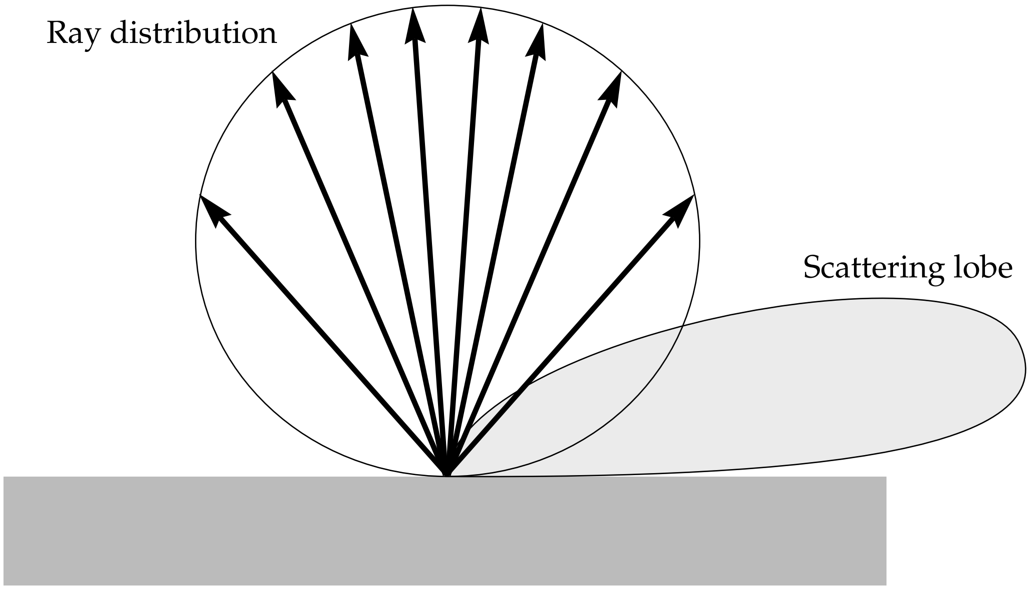

Figure A.2: Problems with $\cos\widehat{\theta}_r$ Sampling

local zenith and fewer toward the horizon. Unfortunately, directional-diffuse scattering works badly with this distribution (Figure A.2); as $\widehat{\theta}_i \to \frac{\pi}{2}$, the peak of the scattering lobe approaches $\widehat{\theta}_r = \frac{\pi}{2}$ and the shape of $\widehat{\rho}_{bd}$ for most directional-diffuse scatterers tends to grow narrower and more concentrated [He 91]. The cosine distribution thus *avoids* sampling the most important directions: those with maximum contribution to the integral and maximum variation of the function $\widehat{\rho}_{bd}\cos\widehat{\theta}_r$.

A.5 Russian Roulette

Ray tracing can be made much more efficient by terminating the paths of rays whose contribution to the result is insignificant [Hall 83]. In a series of interactions with surfaces, the weight of a ray is typically diminished. Following every ray until it leaves the model entirely could be very expensive with not much contribution to the result. Unfortunately, the usual method of terminating rays introduces *bias*, that is the expected value of the estimator does not equal the value of the integral to be estimated. While bias can sometimes be tolerated to gain some offsetting advantage [Kalos 86, pp. 35-37], it is usually undesirable, and the designer of a Monte Carlo algorithm must be careful to avoid it when possible.

Typical ray tracers terminate weak rays by some method like

```
if ( weight > T )
  c = weight * trace_ray();
else
  c = 0;
return c;
```

If we take

$$C = \frac{1}{N} \sum_{n=1}^{N} c(x) \tag{A.10}$$

as a Monte Carlo estimator for the desired value, then the expected value of C is

$$\begin{aligned} E(C) &= p_a \langle c(x) \rangle + (1 - p_a) \times 0 \\ &\neq \langle c(x) \rangle \end{aligned} \tag{A.11}$$

where p_a is the probability that the weight of a ray will remain above the termination threshold T. This estimator is biased, because its expected value is always less than the true value of the integral. The bias declines as the threshold is reduced and $p_a \to 0$, but this reduced bias comes at the price of increased computation. The ray tracer used in this thesis alters the algorithm slightly to terminate rays with *Russian roulette,* a stochastic technique that removes the bias while retaining efficiency:

```
if ( weight > T )
  c = weight * trace_ray();
else if ( rand() < p ) {
  weight = weight / p;
  c = weight * trace_ray();
}
else
  c = 0;
return c;
```

where `rand()` returns random values from a uniform distribution between 0 and 1, and `p` is the (constant) probability p_c that a ray whose weight falls below the threshold `T` will survive. The expected value then becomes

$$\begin{aligned} E(G) &= p_a \langle c(x) \rangle + (1 - p_a) \left[\langle c(x) \rangle \, p_c \frac{1}{p_c} + 0 \times (1 - p_c) \right] \\ &= \langle c(x) \rangle \, . \end{aligned} \tag{A.12}$$

This creates an unbiased estimator regardless of the values of `T` and `p`; increasing either to gain speed increases variance, but not bias.

A.6 Stratification

An elementary variance reduction technique is stratification (Section 4.4). The ray tracer described here stratifies in six dimensions: the incident angles (θ_i, ϕ_i), the (x, y) location of the initial target point on the notional plane, and the reflection angles $(\widehat{\theta}_r, \widehat{\phi}_r)$ at each diffuse scattering event. The incident and reflection hemispheres are stratified in a $n \times 4n$ rectangular pattern in $(\cos\widehat{\theta}_r, \widehat{\phi}_r)$; this gives strata of equal solid angle.

A.7 Converting the Ray Tracer for Rendering

All the details above apply to any geometric optics simulation: not only in tracing rays to predict the BRDF, but also in energy-consistent rendering. The ray tracer used to obtain BRDFs was also used, with slight modifications, to generate images with correct global illumination for this thesis. The main changes are:

- Choosing rays diverging from the viewpoint and distributed over the image plane, rather than choosing rays from random directions over the incident hemisphere;
- Removing the estimator for spherical harmonics coefficients;
- Converting energy flux density E to radiance I for the eye rays; and
- Sampling emitters directly to reduce variance.

The first change simply replaces the ray selection process with that from a traditional ray tracer used for rendering. The second also makes the ray tracer more "conventional" in that the quantity estimated is an array of pixels, rather than a matrix of coefficients for the BRDF. The third is necessary because this ray tracer calculates energy flux density, rather than radiance, as ray tracers since [Goldstein 71] have traditionally done. The conversion is

$$dI = \frac{dE}{\cos \widehat{\theta}_i d\omega_i} \tag{A.13}$$

where $\widehat{\theta}_i$ is the elevation angle at which the eye ray strikes the first surface, and $d\omega_i$ is the solid angle subtended by the pixel.

Sampling emitters is a simple case of importance sampling. Since emitters are generally much brighter than the rest of the environment, their contribution to the irradiance at a point will usually be large. We can reduce variance by sampling the emitters at a greater density than the rest of the environment. The following steps accomplish this.

- Find a bounding sphere for each emitter.
- At each receiving point, sample the cone formed by the point and the bounding sphere for the emitter.
- whenever a ray intersects an emitter without being occluded, add in a contribution from the emitter.
- Normalize the sum of contributions according to the proportion of the hemisphere subtended by the solid angle of the sampling cone.

The emitter cone is sampled in a radially-stratified pattern to reduce variance further.

The entire incident sphere is sampled separately for indirect illumination, discarding any samples that directly hit an emitter. Two emitters, or their sampling cones, may overlap for a given receiving point; this can result in hits on one emitter while sampling another. We ignore these hits, counting them as misses, as they would cause a given emitter to be sampled twice. Samples can thus be divided into two classes, those that contribute to the pixel value:

- Emitter samples that strike the intended emitter, and
- Indirect samples of the hemisphere that strike a non-emitter;

and those that are discarded:

- Emitter samples that miss the chosen emitter entirely (including those that strike another emitter),
- Emitter samples that strike an occluding object (possibly another emitter) before striking the chosen emitter, and
- Indirect samples of the hemisphere that strike an emitter.

Bibliography

[Appel 68] Appel, Arthur. Some techniques for shading machine renderings of solids. In *Proceedings of 1968 Spring Joint Computer Conference* (Atlantic City, New Jersey, April 30-May 2, 1968) *AFIPS Conference Proceedings* 32, 37-45.

[Arvo 90] Arvo, James, and Kirk, David. Particle transport and image synthesis. In *Proceedings of SIGGRAPH '90* (Dallas, Texas, August 6-10, 1990) *Computer Graphics* 24(4) (August 1990), 63-66.

[Blinn 77] Blinn, J., Models of light reflection for computer synthesized pictures. In *Proceedings of SIGGRAPH '77* (Boulder, Colorado, July 27-31, 1977) *Computer Graphics* 11(2) (July 1977), 192-198.

[Cabral 87] Cabral, Brian, Nelson. Max, and Springmeyer, Rebecca. Bidirectional reflection functions from surface bump maps. In *Proceedings of SIGGRAPH '87* (Anaheim, California, July 27-31, 1987) *Computer Graphics* 21(4) (July 1990), 273-281.

[Cook 82] Cook, R. L, and Torrance, K. E. A reflectance model for computer graphics. *ACM Transactions on Graphics* 1(1) 1982, 7-24.

[Cook 84] Cook, Robert L, Porter, Thomas and Carpenter, Loren. Distributed ray tracing. In *Proceedings of SIGGRAPH '84* (Minneapolis, Minnesota, July 23-27, 1984) *Computer Graphics* 18(3) (July 1984), 137-145.

[Gouraud 71] Gouraud, H. Continuous shading for curved surfaces. *IEEE Transactions on Computers* 20(6) (June 1971), 623-628.

[Goldstein 71] Goldstein, Robert A., and Nagel, Roger. 3-D visual simulation. *Simulation* 16(1) (January 1971), 25-31.

[Goel 91] Goel, Narendra S., Rozehnal, Ivan and Thompson, Richard L. A computer graphics based model for scattering from objects of arbitrary shapes in the optical region. *Remote Sensing of Environment* 36(2) (1991), 73-104.

[Hall 83] Hall, Roy A. *A Methodology For Realistic Image Synthesis.* Master's Thesis, Cornell University, 1983.

[Hecht 87] Hecht, Eugene. *Optics.* Second Edition, Addison-Wesley, 1987.

[He 91] He, Xiao D., Torrance, Kenneth E., Sillion, François X., and Greenberg, Donald P. A comprehensive physical model for light reflection. In *Proceedings of SIGGRAPH '91* (Las Vegas, Nevada, July 28-August 2, 1991) *Computer Graphics* 25(4) (July 1991), 175-186.

[Goral 84] Goral, Cindy M., Torrance, Kenneth E., Greenberg, Donald P., and Battaile, Bennett. Modeling the interaction of light between diffuse surfaces. In *Proceedings of SIGGRAPH '84* (Minneapolis, Minnesota, July 23-27, 1984) *Computer Graphics* 18(3) (July 1984), 213-222.

[Howell 63] Howell, John R., and Perlmutter, Morris. Directional behavior of emitted and reflected radiant energy from a specular, gray, asymmetric groove. *NASA Technical Note* D-1874, NASA Lewis Research Center, Cleveland, Ohio, August, 1963.

[Jones 85] Jones, M. N. *Spherical Harmonics and Tensors for Classical Field Theory.* John Wiley & Sons, New York, 1985.

[Kajiya 89] Kajiya, James T., and Kay, Timothy L. Rendering fur with three dimensional textures. In *Proceedings of SIGGRAPH '89* (Boston. Massachusetts, July 31-August 4, 1989) *Computer Graphics* 23(3) (July 1989), 271-280.

[Kajiya 86] Kajiya, James T. The rendering equation. In *Proceedings of SIGGRAPH '86* (Dallas, Texas, August 18-22, 1986) *Computer Graphics* 20(4) (August 1986), 143-150.

[Kajiya 85] Kajiya, James T. Anisotropic reflectance models. In *Proceedings of SIGGRAPH '85* (San Francisco, California, July 22-26, 1985) *Computer Graphics* 19(4) (July 1985), 15-21.

[Kajiya 84] Kajiya, James T., and Von Herzen, Brian J. Ray tracing volume densities. In *Proceedings of SIGGRAPH '84* (Minneapolis, Minnesota, July 23-27, 1984) *Computer Graphics* 18(3) (July 1984), 165-174.

[Kalos 86] Kalos, Malvin H., and Whitlock, Paula A. *Monte Carlo Methods*. John Wiley & Sons, 1986.

[Oppenheim 89] Oppenheim, Allan V. and Shafer, Ronald W. *Discrete-Time Signal Processing.* Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

[Nicodemus 77] Nicodemus, F. E., Richmond, J. C., Hsia, J.J., Ginsberg, I. W., and Limperis T. Geometrical considerations and nomenclature for reflectance. *National Bureau of Standards Monograph* 160, October 1977.

[Phong 75] Phong, Bui-Tuong. Illumination for computer generated pictures. *Communications of the ACM* 18(6) (June 1975), 311-317.

[Polgar 65] Polgar, Leslie G., and Howell, John R. Directional thermal-radiative properties of conical cavities. *NASA Technical Note* D-2904, NASA Lewis Research Center, Cleveland, Ohio, June, 1965.

[Poulin 90] Poulin, Pierre, and Fournier, Alain. A model for anisotropic reflection. In *Proceedings of SIGGRAPH '90* (Dallas, Texas, August 6-10, 1990) *Computer Graphics* 24(4) (August 1990), 273-282.

[Shirley92] Shirley, Peter. Nonuniform Random Point Sets via Warping. in *Graphic Gems III.* Kirk, David, ed. Academic Press, New York, 1992, 80-83.

[Sillion 91] Sillion, François X., Arvo, James, Westin, Stephen H., and Greenberg, Donald P. A global illumination solution for general reflectance distributions. In *Proceedings of SIGGRAPH '91* (Las Vegas, Nevada, July 28-August 2, 1991) *Computer Graphics* 25(4) (July 1991), 187-196.

[Siegel 81] Siegel, R., and Howell, J. R. *Thermal Radiation Heat Transfer.* Hemisphere Publishing, New York, 1981.

[Torrance 67] Torrance, Kenneth E. and Sparrow, Ephraim M. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57(9), September 1967, 1105-1114.

[Torrance 66] Torrance, Kenneth E. and Sparrow, Ephraim M. Off-specular peaks in the directional distribution of reflected thermal radiation. *Journal of Heat Transfer*, May 1966, 223-230.

[Wallace 84] Wallace, P. R. *Mathematical Analysis of Physical Problems*. Dover Publications, Mineola, N. Y. 1984.

[Ward 88] Ward, Gregory J., Rubinstein, Francis M., and Clear, Robert D. A ray tracing solution for diffuse interreflection. In *Proceedings of SIGGRAPH '88* (Atlanta, Georgia, August 1-5, 1988) *Computer Graphics* 22(4) (August 1988), 85-92.

[Ward 92] Ward, Gregory J. Measuring and modeling anisotropic reflection. In *Proceedings of SIGGRAPH '92* (Chicago, Illinois, July 27-31, 1992) *Computer Graphics* 26(4) (July 1992), 265-272.

[Whitted 80] Whitted, Turner. An improved illumination model for shaded display. *Communications of the ACM* 23(6) June 1980, 343-349.

[Wylie 67] Wylie, C., Romney, G., Evans, D., Erdahl, A. "Half-tone Perspective Drawings by Computer. In *Proceedings of 1967 Fall Joint Computer Conference* (Anaheim, California, November 14-16, 1967) *AFIPS Conference Proceedings* 31, 49-58.